AD-A038 978 UTAH UNIV SALT LAKE CITY DEPT OF COMPUTER SCIENCE F/G 6/16

MONAURAL SENSITIVITY TO DISPERSION IN IMPULSES AND SPEECH.(U)

DEC 75 W H GREER DAHC15-73-C-0363

UNCLASSIFIED UTEC-CSC-75-142 NL

1 OF 2

AD
A038978

Monaural sensitivity to dispersion in impulses and speech

6 B.S.

WILLIAM HARVEY GREER
UNIVERSITY OF UTAH

See 1473 in back

DECEMBER [illegible]
UTEC-CSc-[illegible]
COMPUTER SCIENCE, UNIVERSITY OF UTAH
SALT LAKE CITY, UTAH [illegible]

MONAURAL SENSITIVITY TO DISPERSION IN IMPULSES AND SPEECH

by

William Harvey Greer

December 1975 UTEC-CSc-75-142

This research was supported by the Advanced Research Projects Agency of the Department of Defense, under contracts DAHC15-73-C-0363 and F30602-70-C-0300.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Little is known of the ear's phase sensitivity limits outside the well known fact that excessive dispersion in speech signals results in chirp-like sound quality. The limits of phase sensitivity may be determined in reference to the parameters of an idealized dispersion since phase shifts are nothing more than delays in the short-time spectra of a signal. A number of discrimination tests were conducted which contrasted standard stimuli with phase modified variants of the stimuli to determine the limits of phase sensitivity. The idealized dispersion was that of delaying a band of frequency components such that discriminability could be measured as functions of center frequency, bandwidth and delay. The stimuli used in the tests included an impulse, phonemes, words, and a sentence.

Sensitivity to dispersion in impulses was shown to be dependent on intensity, center frequency, bandwidth, and delay. Discrimination scores half way between chance and perfect performance were achieved for dispersed impulses with delayed bands of frequency components centered between 250 and 500 Hertz and

---

This report reproduces a dissertation of the same title submitted to the Department of Communication, University of Utah, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

delayed between 0.0625 and 0.125 milliseconds. Sensation level of the stimulus pairs was 80 dB. The effect of bandwidth on discriminability appeares to reach a plateau at 100 Hertz, which is to say that bandwidths of interest are less than 100 Hertz.

The neutral vowel, the plosive /t/, and the fricative /f/ were used as stimuli in experiments in which phonemes served as the standard stimuli. The stimulus pairs were presented at levels approximating normal conversational level speech, i.e., 70, 32, and 35 dB Sensation Level respectively. Discriminability was observed to be strongly dependent on bandwidth and delay. A very slight, if any, decrease in discriminability was observed with increased center frequency. Rank ordering of the phonemes in decreasing sensitivity to dispersion is plosive, vowel, and fricative. For bandwidths of approximately 100 to 400 Hertz, the plosive requires 4 to 8, the vowel 8 to 16, and the fricative 16 to 32 milliseconds of delay. For wider bandwidths, the respective ranges are 2 to 4, 2 to 4, and 4 to 8 milliseconds. More explicitly, for narrower bandwidths, the plosive requires 4 to 8 milliseconds for 100 to 200 Hertz bandwidth, the vowel 8 to 16 milliseconds for 100 to 400 Hertz bandwidth, and the fricative 16 to 32 milliseconds for 100 to 400 Hertz bandwidth. For wider bandwidths the plosive requires 2 to 4 milliseconds for 200 to 400 Hertz bandwidth, the vowel 2 to 4 milliseconds for 200 to 300 Hertz bandwidth, and the fricative 4 to 8 milliseconds for 400 to 800 Hertz bandwidth. Less delay is required for greater

bandwidth since discriminability is directly dependent on both independent variables.

Discriminability scores obtained from tests using words and a sentence as standard stimuli agree with the scores obtained from tests using phonemes. Discriminability half way between chance and perfect performance was obtained for the dispersed sentence in which a 400 Hertz band of components centered at 500 Hertz was delayed between 4 and 8 milliseconds. The score agrees with that for a vowel which implies that cues to dispersed sound quality in continuous speech arises from vowel dispersions.

The results of the study indicates that speech processing systems introducing no more than a few milliseconds of dispersion will cause little detriment to the speech quality. This conclusion is significant inasmuch as it is difficult, if not impossible, to determine the phase of many speech processing systems.

CHAPTER I

INTRODUCTION

Phase Sensitivity and Psychophysics

Early investigators of hearing science had long understood that the pitch and loudness of a sinusoidal sound are correlated respectively with the frequency and intensity of the sound source. The physical correlates of sound quality, however, were not as well understood. Helmholtz (1863) showed that the only attribute of a periodic sound to which quality could be correlated was the harmonic composition of the sound.

Helmholtz satisfied himself that the correlates of most importance to sound quality were the period and amplitude of the constituent frequency components and that phase was of little consequence. Research since the time of Helmholtz has shown that the ear is undoubtedly sensitive to phase. The full nature of phase sensitivity, however, remains incompletely understood notwithstanding the importance of understanding phase perception for both theoretical and practical reasons.

The theoretical study of psychophysics as it pertains to phase sensitivity, or any other aspect of perception, as a facet of man's communicative system is justified in that man himself is an interesting subject of study. As a consequence of such interest, a goal of the current study is to provide a broader theoretical

insight into the nature of phase sensitivity.

The Role of Psychophysics in the Design of Speech Processing Systems

The study of phase sensitivity is justified for practical reasons because of the role played by the ear as the final link in speech processing systems. Speech processing systems might include those for coding and decoding, transmission, storage and retrieval, restoration, enhancement, etc. Restoration systems would be employed to retrieve signals from excessive noise or other contaminations. Signal enhancement processing could be included in the design of hearing aids and communication systems.

The design of speech processing systems must be approached with an awareness of the capabilities and limitations of all the individual components, including the ear, which constitute the system. For example, knowing that the average ear is capable of correctly interpreting words transmitted through a transmission system of only 3000 Hertz bandwidth is extremely important in the establishment of reliable and low cost commercial telephony. Knowledge of the ear's capabilities dictates a greater transmission bandwidth and lower permissible distortion levels for the fidelity transmission of music in home entertainment systems.

Because of the real need for specific psychophysical data applicable to speech processing system design, the major goal of the current study is to provide more pertinent data pertaining to phase sensitivity.

Computer Usage in Psychophysical Research

In his review of hearing science, Boring (1938) noted that there were few facts known concerning hearing when Helmholtz wrote his _Lehre von den Tonempfindungen_ in 1863. Most of the facts were not very old. Although the science of sound may be dated back to Pythagoras, Boring noted that it was not until the advent of the electronic vacuum tube by de Forest in 1907 that the science of sound became more a field of success than effort. The major contribution of the vacuum tube was to provide accuracy of measurement in psychophysical experimentation. The adoption of electronics to the science of hearing however, was not fully realized until very late in the 1920's.

A new era in the development of hearing science began with the rise of modern digital computers in the 1950's. The power, precision, and versatility of the digital computer holds great promise in the further advancement of psychophysical research. The major contribution of computers will result from their application to the solution of exceedingly complex problems. Indeed, because of the complexity of the problems that can now be practically approached, many modern developments in psychophysics would be impossible without the aid of the large and powerful digital processing machines.

The digital processing of signals represents a major use of computers in psychophysical research. One salient advantage of digital signal processing is that the process need not be linear. Examples of such research include the pioneering work of Stockham

(1972, 1975) in which he was able to remove the convolutional effects of the recording horn from old Caruso sound recordings using the principles of homomorphic filtering, i.e., generalized linear filtering (Oppenheim *et al.*, 1968). Miller (1973) refined and digitally implemented homomorphic vocoder techniques for the restoration of old, noisy sound recordings. The nonlinear processing required in the restoration was made practical only by the power of the digital computer.

Other powerful and interesting applications of computers in psychophysics include simulation, stimulus preparation, control of experiments, data acquisition, and data reduction. As an example of simulation, Callahan (1975), using a two-dimensional speech processing algorithm, has developed the capability of modeling auditory effects such as recruitment, fatigue, adaptation, inhibition, etc.

A major thrust of the research to be described here has been to further develop and apply digital techniques which may be generally applicable to psychophysical research. It is important not to allow one's thinking to be overly constrained by the limitations of current computer architecture when developing digital techniques. The advance of computer science is so rapid that larger machines become available before the potential and capacity of machines already in use may be fully explored and exhausted. Much understanding of the practical application of digital techniques to future psychophysical research may be lost if the full potential of present day machines is not fully investigated. The study of the

computer as a basic research tool is justified in its own right inasmuch as the proven usefulness of digital computers has made computer application to psychophysical research as basic as psychometric methods.

Uniqueness and Significance of the Study

This study differs from previous studies concerning the psychophysical determination of phase sensitivity in three fundamental respects. First, the philosophical approach to the problem is different. Previous studies, using a variety of stimuli and psychophysical methods, have indirectly attempted to "prove" or "disprove" Helmholtz' phase rule by demonstrating new "phase effects." The results of those studies, from the second half of the nineteenth century to date, have been conflicting. Earlier failures in demonstrating phase sensitivity are probably due to equipment limitations. Any successful demonstration of phase sensitivity was attributed to experimental artifacts as discussed by Beasley (1930a). One must be impressed by the fact, however, that since the second quarter of the current century almost all phase sensitivity studies, without exception, have demonstrated the ear's sensitivity to phase modifications of one type or another. Notwithstanding the success of the later experiments, there has yet to be a critical study unifying the results into a comprehensive theory of phase sensitivity.

The approach taken by the author in this study is neither to prove or disprove the phase rule or to demonstrate a new phase effect. It will be clear from the review of literature in Chapter

III that the ear is sensitive to phase. The phase effects demonstrated by earlier investigators have generally relied on specialized stimuli not normally encountered in the every day acoustic environment. The author, accepting the sensitivity of the ear to phase, proposes to measure the sensitivity by systematically modifying the phase of impulses and speech sounds. The results of the study may be more pragmatic than many of the other studies discussed in the review of literature but none the less applicable to theoretical insights.

The second respect in which the present study differs from previous ones is in the computer instrumentation of the study. It is believed by the author that the computer has been put to fairly sophisticated use during the course of the investigation to be discussed. The uses to which the computer has been put will be discussed in greater detail throughout this report. It is hoped that the discussion will be of profit to future psychophysical experimentation of the type described herein.

The third respect in which the study differs is in the further development of the computer as a basic research tool.

The significance of the study is inextricably involved with it's uniqueness. The determination of phase sensitivity has been shown to be interesting for both theoretical and practical considerations. The very least application to which the results may be applied is in the "worse case" design of signal processing systems. The instrumentation of the study and the development of digital techniques are of similar significance.

In summary, a study is to be presented pertaining to the phase sensitivity of the ear. The role of psychophysics in the design of speech processing systems and the use of computers in psychophysical research has been discussed. The present study is described as being unique because of its approach to the phase sensitivity problem, its instrumentation, and its development of digital techniques.

CHAPTER II

THE PHASE SENSITIVITY QUESTION

Acoustic signals may be analyzed by Fourier's methods into their constituent frequency components. Such analysis may be effected mathematically, mechanically, optically, electrically, and digitally. That the ear may also effect frequency analysis after the manner of Fourier was first stated by Ohm in 1843. Helmholtz (1863) verified the law and as a result of his own investigation added the generally accepted dictum that the musical quality of a musical tone is phase independent. The ear does appear, however, to be sensitive to phase modifications of certain types. Determination of the way and to what extent the ear is sensitive to phase is termed by the author to be the phase sensitivity question. There has been much work since the time of Helmholtz concerned with determining all the varied ramifications of the phase sensitivity question. The following discussion is expedient to clarify the phase sensitivity question and to put the current work into its proper perspective.

The Time and Frequency Domain Representation of Acoustic Signals

Acoustic signals are the spherical propagation of acoustic energy from a given source. The information content of such signals is represented as atmospheric pressure variations as measured at

some fixed point from the source. A graphical representation in which the pressure or velocity variations are plotted as a function of time is known as the <u>signal waveform</u>.

Waveforms may be described in terms of mathematical functions. It was generally known before the time of Fourier[1] that continuous functions could be represented as the sum of sinusoids of various periods, amplitudes, and phase relationships, but it remained for Fourier to demonstrate that discontinuous waveforms could also be so represented.[2]

It follows from the preceding that a function may be represented as either a waveform in the so called <u>time domain</u> or as the weighted sum of sinusoids in the so called <u>frequency</u> or <u>Fourier domain</u>. Changing from one representation to the other is termed a <u>transformation</u>; in particular, changing from the time domain to the frequency domain is termed the <u>Fourier transformation</u>, the inverse process being termed the <u>inverse Fourier transformation</u>. Transformations are performed in order to exploit the properties of one domain or the other. Common examples of transformational exploitation are found in statistics, the use of logarithms, etc. The particular properties of the Fourier transformation of interest to this study will be discussed in Chapter IV, but the specific purpose of the present introduction is to provide a historical

---

[1]*Jean Baptiste Joseph Fourier* (1768 - 1830), a distinguished French mathematician.

[2]Guillemin (1963), has noted that Euler and contemporary mathematicians felt that such representation was impossible except for continuous functions. Although Fourier did not offer rigorous proof, such proof was demonstrated by Dirichlet in 1837.

perspective for Ohm's contribution to the science of hearing.

Ohm's Law of Acoustics

The essence of Ohm's law of acoustics is that the ear performs a Fourier analysis on an acoustic signal by transforming the signal into sinusoidal components which the ear individually perceives. Ohm (1843) developed the law in part from the 1841 experimental work of Seebeck, who Wightman and Green (1974) cite as probably being the first to conduct a systematic investigation of pitch perception. Through a series of experiments involving the use of a siren, Seebeck concluded that the pitch of a sound was related to the periodicity of the waveform rather than the spectral energy at the reciprocal of the periodicity of the waveform, i.e., the fundamental frequency.

Ohm took issue with Seebeck's conclusion in that Ohm believed that a pitch could be heard only if the acoustic signal contained power at that frequency. Ohm demonstrated by use of the Fourier transformation that Seebeck's waveform did in fact contain the required frequency components, but at low intensities. Seebeck, in turn, objected to Ohm's explanation because the amplitude of the components were too low to account adequately for the intensity of the pitch which could be heard. The controversy was tentatively resolved twenty years later by Helmholtz, who postulated that nonlinear distortion in the middle ear would produce distortion products that would combine with the incidental signal so as to be analyzable by the ear into component frequencies as explained by Ohm's law.

Verification of Ohm's Law of Acoustics

Helmholtz (1863), as part of his investigation into the physics and psychophysiology of music, had occasion to determine whether or not the ear analyzed an acoustic signal or not [p. 33], and reasoned that inasmuch as the ear is able to analyze an acoustic signal produced by two tuning forks that it must also be able to so analyze a note produced by a single instrument such as a flute or organ pipe. Helmholtz cited Ohm as having first laid down the rule by which the ear analyzes acoustic signals and according to his understanding of the rule, stated that

> Every motion of the air, then, which corresponds to a composite mass of musical tones, is, according to Ohm's law, capable of being analyzed into a sum of simple pendular vibrations, and to each such single simple vibration corresponds a simple tone, sensible to the ear, and having a pitch determined by the periodic time of the corresponding motion of the air.

Having so restated the law, Helmholtz [p. 52] proceeded to prove it by the expedient of showing that the ear does not perceive frequency components when they are not part of an acoustic signal. Whether or not the components were present in a signal was determined either by mathematical calculation or by "sympathetic resonance," i.e., by the use of tuned resonators which would amplify the given component in the signal such that it would be audible. A signal produced by a vibrating string was used because the constituent frequencies of the sound could be easily changed by the manner and the spot in which the string was excited, and because of the ease in which the theoretical or experimental analysis of the sound could be effected.

Helmholtz used the principle of sympathetic vibration of other strings and resonators for the experimental analysis of a vibrating string. He also used the more direct approach of touching the string at its various vibrational nodes and observing which components disappeared. The reader is referred directly to Helmholtz [p. 52] for a fuller appreciation of the simplicity and elegance in which the frequency component composition of a vibrating string may be determined simply by touching the string at its vibrational nodes.

By the expedient of damping vibrating strings at various points, thereby removing the corresponding frequency components, and then demonstrating that the ear could no longer hear that particular frequency component, Helmholtz demonstrated that the ear analyzed the sound into the same components as would be analyzed by physical analysis. Notwithstanding the inability of the ear to recognize among the strong components of a signal all the components detectable with the resonators, Helmholtz concluded his proof with the statement [p. 56]

> The ear recognizes without resonators the simple tones [sinusoidal components] which the resonators greatly reinforce, and perceives no upper partial tone [i.e., other frequency components] which the resonator does not indicate. To verify this conclusion, I performed numerous experiments, both with the human voice and the harmonium, and they all confirmed it.

Helmholtz not only supported Ohm's law but also provided a physiological basis for its operation. The basilar membrane of the middle ear, based on the anatomical discoveries of Corti in 1851, was postulated to be composed of a series of transversely stretched

fibers, each of which was resonant to a different frequency. An acoustic signal vibrating the membrane would consequently excite only those fibers which were tuned to the resonant frequency components in the signal. Corresponding individual sensations would follow from the doctrine of specific nerve energies stated by Muller in 1836.

The Sound Quality Question

Helmholtz was interested in determining why different musical instruments (including the human voice) had different sound qualities, e.g., that peculiar property of sound that enables a listener to recognize its source. As introduced in Chapter I, Helmholtz showed that the only attribute of a sound to which quality could be *correlated was the shape of its representative waveform.* Helmholtz [p. 65] admitted that the reason for the conclusion was negative but proceeding from the results discussed in the last section concluded that musical tones of the same quality would always be composed of the same frequency components inasmuch as it is the components that elicit their corresponding sensations. The natural question following the conclusion is to what extent the difference in quality can be explained by the combination of different components of various amplitudes.

Helmholtz was able to demonstrate through a series of experiments that *sound quality was correlated with* the weighted sum of frequency components constituting the sound. Helmholtz found it important to note that quality is not to be confused with the peculiarities of how the sound begins or ends. Sounds produced by

musical instruments may build up and die away at different rates, etc. Helmholtz further noted that even when a musical tone is uniformly sustained that certain noises may accompany it such as the hissing of air in wind instruments, the rubbing of a violin bow, etc. In light of these considerations, Helmholtz [p. 67] defined musical quality as follows.

> When we speak in what follows of musical quality of tone, we shall disregard these peculiarities of beginning and ending, and confine our attention to the peculiarities of the musical tone which continues uniformly.

The Phase Rule

After investigating the musical quality of various musical instruments and the human voice, Helmholtz [p. 119] turned his attention toward determining the importance of the phase relationships among the constituent frequency components to sound quality. It should be noted that the sounds under investigation were musical tones, i.e., sounds which are periodic as opposed to irregular motions of the air such as noise. Helmholtz pursued his goal by using various synthesized musical tones which imitated vowels to determine whether or not a difference in quality was detectable when the phase was varied.

A battery of electrically driven tuning forks was used in the synthesis. Phase could be altered by bringing the resonance chamber of a given fork slightly out of tune with the fork. Every phase condition was possible using this technique. Placing the chambers out of resonance also weakened the sound of the fork but such weakening was compensated by adjusting the distance between the

other forks and their corresponding resonators. The same effect could be achieved by slightly mistuneing the forks, e.g., by adding drops of sealing wax to the tines. In such cases, the actual phases could be measured by viewing Lissajous figures by means of a vibration microscope [p. 126].

Helmholtz experimented with many tones with different phases and was never able to experience any difference in the quality of the tone. He found that it made no difference whether he weakened the constituent frequency components by detuning their resonant cavities, thereby shifting their phase, or by moving the resonators further from their forks. Helmholtz [p. 126] laid down the following rule as a result of these experiments:

> . . . the quality of the musical portion of a compound tone depends solely on the number and relative strength of its partial simple tones, *and in no respect on their differences of phase.*

Helmholtz qualified his rule, however, with the following restriction [p. 127]:

> It must be here observed that we are speaking only of musical quality as previously defined. When the musical tone is accompanied by unmusical noises, such as jarring, scratching, soughing, whizzing, hissing, these motions are either not to be considered as periodic at all, or else correspond to high upper partials, of nearly the same pitch, which consequently form strident dissonances. We were not able to embrace these in our experiments, and hence we must leave it for the present doubtful whether in such dissonating tones difference of phase is an element of importance. Subsequent theoretic considerations will lead us to suppose that it really is.

The Validity of Ohm's Law

The validity of Ohm's law has been discussed since its first publication. It becomes apparent, however, in reviewing the

literature concerning the law that there are differences in opinion as to what constitutes the law [See Appendix A]. It is generally accepted that the law pertains to the frequency analyzing ability of the ear. Confusion arises as to whether or not Ohm included statements in his law pertaining to phase sensitivity or if such statements were later added by Helmholtz. A careful reading of Ohm and Helmholtz reveals that Ohm did not conduct any experiments on phase or make any statements concerning phase sensitivity. Such statements were later added by Helmholtz as a result of his own work. Goldstein (1967) gives the interpretation that Helmholtz' phase rule complements Ohm's law in that the spectral components one may "hear out" due to the Fourier analysis are phase independent.

This report is mainly concerned with the phase sensitivity question. It has been necessary, however, to consider the phase question up to this point in conjunction with Ohm's law for two reasons. The first reason is that it was probably Helmholtz who was first to make a definitive statement about phase sensitivity. To the extent that Helmholtz demonstrated the correlation between the musical quality of a tone and the weighted sum of its constituent frequency components, and that the ear analyzed the sound according to the tenets of Ohm's law, the work of Ohm and Helmholtz are inseparably connected. The second reason for the joint consideration of Ohm's analytic law and Helmholtz' phase rule is the previously discussed confounding to the two.

Exceptions to Ohm's law as defined in its restricted sense are not directly pertinent to the thesis under consideration but

because of the long time intimacy between the work of Ohm and Helmholtz, it must be mentioned for completeness that there appear to be minor exceptions to Ohm's law.

Richardson (1927) has indicated that Ohm's law holds fairly well for weak sounds and that "the instances where it seems to be untrue can be explained in the main as aural illusions, that is to say, that their cause is psychological." Richardson stated that "this is not a refutation of Ohm's law, at least in principle, but may simply imply the intrusion of other simple tones not in the external sound, into the quality of the note as perceived by the ear." Richardson cited subjective combination tones as an example of such intrusions.

Trimmer and Firestone (1937) noted that "these empirical facts--the validity of Ohm's law at low amplitudes and the exceptions at higher amplitudes--are well established . . . ." Stevens and Davis (1938) noted in reference to Ohm's law of acoustics that ". . . the ear is in general able to detect the presence of component frequencies in a sound-wave and to identify their pitch provided they are not too numerous or too faint."

Monaural Phase Effects

It may be generally said that the ear is not indifferent to phase if differences in signal quality is perceived as a result of phase modifications within the signal complex. Such observations of quality change are referred to in the literature as monaural phase effects, and are defined by Trimmer and Firestone (1937) as follows:

> Suppose an observer listens to a combination of n objective tones, all having absolutely fixed frequencies, amplitudes

> and phases. Let Po represent the periodic pressure wave made up of these tones. Let the observer be simultaneously presented with an exploring tone Pe--that is, a tone of which the frequency, the amplitude and the phase are all adjustable:
>
> $$Pe = A \cos (2\pi ft + \theta)$$
>
> If, for given Po, it is possible to find values of A and of f such that the ear hears changes in the combined sound of Po and Pe as θ changes, a (monaural) <u>phase effect</u> is said to be observed. These changes in the heard sound may be of loudness, pitch or quality.

These authors stated that the monaural phase effect is not to be confused with the binaural phase effect which is important in sound localization.

It is possible to generalize the concept above to cover instances in which the "exploring tone" is implicit in the signal complex rather then existing as an explicit entity as described in the definition above.

This definition of a monaural phase effect appears to be commonly accepted in the literature. Stevens and Davis (1938) [p. 203] stated that "those experiments in which an auxiliary tone was made to beat with an aural harmonic prove definitely that the phase-relations among the harmonic components of a stimulus are detectable, for otherwise these beats could not occur," i.e., the perception of beats is a phase effect. It is not clear that only a single mechanism accounts for the detectability of all phase effects. Helmholtz [p. 127] apparently did not intend to use a definition of phase sensitivity as broad as that defined by monaural phase effects, for in reference to his phase rule, he wrote:

> An apparent exception to this rule must here be mentioned. If the forks B♭ and b♭ are not perfectly tuned as Octaves, and are brought into vibration by rubbing or striking, an

attentive ear will observe very weak beats which appear like small changes in the strength of the tone and its quality. These beats are certainly connected with the successive entrance of the vibrating forks on varying difference of phase. Their explanation will be given when combinational tones are considered, and it will then be shewn that these slight variations of quality are referable to changes in the strength of one of the simple tones.

Beasley (1931) noted that Helmholtz considered "musical quality" to mean "vowel quality" and that Helmholtz did not consider relative changes in loudness to be variations in "musical quality." As a further example, perhaps, of not wishing to confound discrimination of phase shifts and discrimination of amplitude changes, even though the amplitude changes result from phase modifications, Hansen and Madsen (1974) took care in their study of phase sensitivity to use a stimulus in which the amplitude changes were held constant.

Thompson's (1877) discovery that beats could be detected when individual tones were presented separately to each ear may be possible justification for restricting the definition of monaural phase effects to exclude amplitude changes. Rayleigh (1907) repeated Thompson's experiment and received the same results. It is not entirely clear that such restrictions should be placed on the definition of monaural phase effects but one should be aware of the possibility of different classes of phase effects and different mechanisms to account for them. The definition of phase deafness appears to be subject to the investigator's interpretation.

In declining to classify his "apparent" exception to the phase rule as a phase effect, Helmholtz is said to have introduced a semantic problem. Koenig (1881) [p. 537] objected in that if tone

quality does partly depend on the relative strength of the higher frequency components, and if the relative strength is modified by phase, then the effect of the phase is "actual, and not merely apparent."

It is important to distinguish phase effects from phase sensitivity. Phase effects arise from arbitrary manipulations of phase, e.g., setting the phase of all frequency components to zero. Phase sensitivity on the other hand is to be understood to arise only from phase modifications that may be effected by linear, stationary systems. The reason for the distinction is that the phase modifications used in the study are effected by linear, stationary systems.

The Validity of the Phase Rule

It is not clear that any of the studies to be cited relevant to the phase sensitivity question have been explicitly designed to "prove" or "disprove" Helmholtz' phase rule or its exception. Indeed, to do so, the study would of necessity have to use the same type stimulus used by Helmholtz in formulation of the rule. Some of the studies to be cited have used simple sinusoidal components as the stimulus, however, and have accordingly made reference to the validity of the phase rule or its exception.

Goldstein (1967) noted that based on Helmholtz' "definitions of musical and nonmusical sounds, his inexplicit restrictions on his phase rule, and his concept of limited frequency resolution, it is clear that Ohm's law [i.e. the phase rule] as commonly conceived is valid only for compound tones with relatively large frequency

separations between constituent simple tones, ..."

The idea of "proving" or "disproving" Helmholtz' phase rule may not be a valid concept as the question is commonly approached. As noted by Licklider (1951), Helmholtz never intended his rule to be so generally interpreted. The phase rule as defined by Helmholtz is almost certainly correct. A cogent objection to too heavy a reliance upon its significance, however, is that it pertains only to a very small subset of the day-to-day signals encountered by the average ear, and, inasmuch as phase can be detected as will be discussed in Chapter III, the rule provides little insight into the operation of any phase detecting mechanism of the ear.

Significance of the Phase Sensitivity Question

The question as to whether or not the ear is sensitive to phase is so poorly formed that it is of very little value. The ill-conditioning of the question rests entirely on the generality of the word "phase." A straight forward psychophysical experiment may be conducted to reveal whether or not the ear is able to discriminate between an original acoustic signal and a phase modified version. Proceeding, one may conduct a second study with another particular type of phase modification. It is, in fact, this process that is represented in the review of literature to be cited in Chapter III. Given the process above, one may determine whether or not the ear is sensitive to a particular type of phase modification. The degrees of freedom in adjusting the phase in a given experiment, however, is without bound. Cited studies will show that the ear is conclusively sensitive to phase modifications.

Inasmuch as phase is a general term, one must conclude that if the ear is sensitive to any type of phase modification that it is sensitive to phase. This conclusion, however, is unfortunately of as little value as the question it proposes to answer.

The author submits that, except for its historical value, the validity of the phase rule is of little interest. Rather than determine if the phase rule is valid or not, the proper approach to accessing the phase sensitivity of the ear is to determine in what way and to what extent the ear is sensitive to phase modification within acoustic signals and to determine the nature of the sensations of the sensitivity. This problem may be approached as a study to determine the perceptible and tolerable limits of the ear to phase distortion which is synonymous with phase changes. The author proposes to determine the general limits of phase sensitivity by using impulses and speech sounds as stimuli in a series of discrimination tests. A "general" type of phase distortion based on a stylized analysis of phase distortion as discussed in Chapter V will be used to systematically modify the stimuli such that phase discriminability may be determined as functions of the parameters of the stylized distortion.

CHAPTER III

PERTINENT LITERATURE CONCERNING PHASE SENSITIVITY

The purpose of this chapter is to review relevant phase literature demonstrating that the ear is undoubtedly sensitive to phase. Very early studies concerning the sensitivity of the ear to phase have been reviewed by Beasley (1931).

> Konig, 1881, conducting experiments with a wave siren, concluded that tone quality is changed with phase displacement of a harmonic. Inasmuch as a wave siren does not produce a fidelity reproduction of the waveform cut in the siren disk, however, his results are not conclusive. Hermann, 1894, also using a siren, concluded that phase displacements are irrelevant and criticized Konig's work on the basis that the intensity variations resulted from changes in the wave pattern on the siren disk. Gray, 1899, using tuning forks, concluded that "the single ear can distinguish no difference between two phases of a simple pure tone; neither can the ear distinguish a phase change in a complete harmony." Emile ter Kuile, 1902, using tuning forks to produce tertiary harmonies in which one of the forks was slightly mistuned so as to produce slow beats, produced definite changes in quality. This same type of stimulus, i.e., one using a "floating" or continuous change of phase, but produced by a <u>tone phaser</u>, was used by Beasley in his own experiments. Lindig, 1903, using a "telephone siren," concluded that tone quality was influenced only by beats between identical frequencies introduced by overtones common to different systems in combination. Lloyd and Agnew, 1909, using telephone receivers and special alternating current generators, concluded that quality was not affected by phase. Hartridge and Cosens, 1922, concluded that "change of phase affects the quality of a musical chord if its constituent tones are accompanied by harmonics; but with pure tones, free from harmonics, change of phase does not audibly affect the quality of the mixed tone."

Beasley also cited a number of authors who theorized about phase sensitivity but did not cite their experimental work, if any.

The type of experiment performed to access the sensitivity of the ear to phase may be broadly classified into two types depending on the intensity of the stimulus. The stimulus may be of weak or moderate intensity such that the ear is not overloaded to produce aural nonlinear distortion, or the stimulus may be of sufficient intensity to force the ear into a nonlinear mode of operation. It must be understood at the onset, however, that the general consensus is that phase effects are in fact mediated by the nonlinearities of the ear. The ear is generally modeled as a linear system, driven with low amplitude signals. The model may be too simple. That the ear is nonlinear is well understood; nonlinearity may arise apparently from any part of the auditory system. Research of the type in which the stimulus intensity is certain to produce nonlinear distortion will be discussed first.

Chapin and Firestone (1934), in the attempt to explain masking, difference tones, and certain kinds of beats, demonstrated that when a 108 Hertz signal, strong enough to produce nonlinear distortion, and one of its harmonics were presented to a listener in various relative phase conditions and intensities, that the tone quality and loudness were influenced significantly by the phase relation of the two components. The authors concluded that the ear distorts the input signal so as to constructively or destructively combine distortion products with the original signal.

Lewis and Larsen (1937) demonstrated phase effects with a difference tone of 130 Hertz generated by two frequencies of 390 and 520 Hertz at 70 dB SPL. They measured the intensity of a 130 Hertz

exploring tone, presented in various phases, that was necessary to make the two component signal just noticeably louder, The phase angle at the minimal intensity at which each of the two subjects could detect a change varied for each subject by 120 degrees. When the minimum phase for each subject was shifted by 180 degrees each subject reported the sound as being at its softest.

When the ear is forced into a nonlinear region of operation by a test stimulus of excessive intensity, the stimulus quality will vary with the onset, growth, and phase relations of the distortion products. Although work in this area will likely lead to further understanding of the auditory system and further revision of the theories of hearing, it is not clear that data gathered from such experiments in which the ear is operating in an abnormal mode will be directly applicable to discovering any basic phase sensitivity mechanism of the ear.

Research of the type involving lower level stimuli has both supported and refuted the phase rule. The view that tone quality is independent of the phase relationships among the frequency components of a sound had apparently become accepted by the time of Beasley's work in the 1930's. Beasley (1930a) objected to this view in that a critical analysis of the experiments that supposedly supported the assertion revealed no evidence either to support or refute the view. In a detailed review of the experimental work of Helmholtz, 1877; Konig, 1881; Hermann, 1894; ter Kuile, 1902; Lindig, 1903; Lloyd and Agnew, 1909; and Hartridge and Cosens, 1921; Beasley reported that a discriminable change in tone quality

invariably occurred in any case in which there was a demonstrable phase shift. Beasley reported that each investigator reviewed suggested that the quality change could be explained by "assuming periodic reinforcement and interference between identical frequencies . . . common to the two generators used as sources for the fundamentals; and that the periodic variation of wave form [sic] consequent upon a change in the phase relations of the fundamentals is irrelevant for hearing."

Beasley objected to the explanation above because the interference effects were deduced from mathematical considerations rather than objective demonstration, because the nature of the variations were not described, and because acceptable proof of phase irrelevance would depend on producing the phase variations under conditions conducive to observing quality variations with no discriminable quality variations.

Beasley reviewed factors necessary to resolve the phase independence question discussed in an earlier paper (Beasley, 1930b). In brief, these factors include the use of test stimuli with maximum waveform variation, due to phase variations between two fundamentals, for a variety of frequency ratios, which occur in a period of time psychologically favorable for observing the change. Under these conditions, the effect of simultaneously shifting the phase between the fundamentals and harmonics of identical frequency and known magnitude should be contrasted with the effects of holding the phase relation between the fundamentals constant while altering the phase relation between harmonics of identical frequency. The

latter test should be done for cases in which the harmonics are separated by various intensity levels, and for cases in which they are of subthreshold magnitude. The last factor is to analyze the effect of cyclically varying the phase between two pure fundamentals.

Beasley determined that there were detectable monaural phase effects with two pure tone stimuli with the frequency ratio 2:3 which were made to vary slowly in their phase relationship. At a signal sensation level of 10 dB the changing phase conditions could not be discriminated any better than chance but at a sensation level between 25 and 30 dB 90 per cent correct discriminations were possible. The subjects reported simultaneous variations in pitch, quality, and pattern. Even though some subjects were able to selectively respond to particular changes they could not agree on what was changing, e.g., the pitch, loudness, etc.

Steinberg (1930) conducted experiments with all-pass systems with monotonically increasing phase characteristics to determine the effect of dispersion on articulation. Steinberg concluded that the major consequence of delay distortion was, in effect, to reduce the band pass of the transmission system by preventing the delayed frequency bands from contributing to articulation.

The experimental work of Schouten (1939) is probably the most recent reporting to support Helmholtz' phase rule and to demonstrate that the qualification to the rule is unnecessary. Schouten described an apparatus used in the synthesis of sound. Paper stencils representing single periods of two waveforms were

placed between a light source and a photoelectric cell. A rotating disk containing radial slits, placed between the stencils and photoelectric cell, was used to vary the light flux incident to the cell as a time function of the desired waveforms. Schouten used this device to investigate, among other things, the influence of phase on sound perception and nonlinear distortion in the ear. As a result of adjusting the relative phase of the two stencils and listening to the combined signal, Schouten reported that "It was never possible to discern any influence of phase on sound perception" except in the special case when nonlinear distortion occurred in the ear.

Schouten pointed out that the device could not be practically applied to shift the phase of a large number of components, notwithstanding the importance of determining the validity of Helmholtz' rule for waveforms with a large number of components. Schouten used four 20-component waveforms with different phases to test phase sensitivity for cases with many components. Schouten stated that "it was found that these four totally different wave forms [sic] were quite indistinguishable as to their sound impression," which he held to confirm Helmholtz' rule for the extreme case of waveforms with many harmonic components. As a practical consequence, Schouten stated that one may confine himself to the measurement of harmonic intensities without regard to phase in the determination of nonlinear distortion.

Mathes and Miller (1947), using modulation techniques to produce essentially a three component signal, determined that the envelope shape of a steady complex signal was important in audible perception. The sensations of roughness or smoothness were found to be influenced by the envelope shape of the waveform. The shape was also found to be related to a sensation of apparent pitch. Differences in sensation could be produced by changing the phase of only a single frequency component or a group of components. The authors reported that their results provide general verification of the qualification placed by Helmholtz on his phase rule. The authors also noted that their results emphasized the importance of time factors in auditory perception.

Flanagan (1950, 1951) studied the effects of delaying or advancing one frequency band relative to the rest of the spectrum on speech articulation and quality. He determined that speech intelligibility was impaired with advances or delays of approximately one-quarter second when the advanced or delayed band was near the center of the spectrum.

Licklider (1957) determined that changes in the phase relations of a 16-component complex signal were discriminable and in some instances sufficient to be of importance in music. It was found that, in general, changing a high-frequency component produced more effect than changing a low-frequency component.

Schroeder (1959) also found a variety of subjective effects as a result of varying the phase relationship in a complex acoustic stimulus. The signal contained up to 31 harmonics. A strong

dependence of the quality on the "peak factor" was found to exist. Some adjustment of phases produced distinct tones which enabled one to play simple melodies.

Craig (1961), and Craig and Jeffress (1962) demonstrated monaural phase effects using stimuli consisting of a 250 Hertz fundamental and its harmonic at various intensities and phase relations. These authors noted that previous studies tended to use complicated stimuli which increased the difficulty of finding a physiological explanation for the phase effects. The authors reported striking individual differences among subjects in their responses. A stimulus which was consistently judged higher in pitch, louder, or purer by one subject was just as consistently judged lower, softer, or less pure by another.

Schroeder (1966) has pointed out that phase is a relatively minor factor in monaurally presented speech signals. There is some influence on speech quality when the sound is presented over earphones but there is probably no effect on intelligibility. Schroeder also noted that phase distortions corresponding to delay distortions exceeding 50 milliseconds will modify the short-time spectrum and will be consequentially detectable as a reverberan speech quality. With sufficiently large delays, the speech can be made unintelligible.

Goldstein (1967) studied the relation of monaural phase perception to limited auditory frequency resolution. Using modulation techniques to produce the stimuli, the author was able to demonstrate that phase effects disappear for stimulus bandwidths

greater than a value proportional to the critical bandwidth of the carrier frequency.

An indirect verification of the sensitivity of the ear to phase has been provided by Allen (1972), who has successfully removed the characteristic buzz from synthetic speech by randomizing the phase of contiguous spectral components.

Hansen and Madsen (1974) showed that the ear was able to detect phase changes without amplitude changes. The authors also noted that phase detectability was increased when loudspeakers having poor transfer characteristics were used as the stimulus transducer.

In summary, a number of studies have been discussed pertinent to the phase sensitivity of the ear with the consensus being that the ear is capable of perceiving phase modifications in acoustic signals.

CHAPTER IV

METHODS AND INSTRUMENTATION

A major thrust of the current research, as introduced in Chapter I, has been the development and application of digital techniques to psychophysical research. Many earlier investigators of the phase sensitivity question had to rely on mechanical, electromechanical, or electrical analog devices with the result that only simple stimuli could be produced and accurately controlled. The computer makes the study under discussion feasible because it makes possible the efficient and accurate generation of the complex stimuli called for in the design of the experiment.

It would be difficult, if not impossible, to use analog equipment to generate the required stimuli. Even if an analog approach were taken, the signals would only be approximate. Approximations would at best introduce errors into the study. The study requires the facility of exactly specifying the phase characteristics of a signal and of systematically modifying them so as to probe a complex hearing phenomenon.

The purpose of the present chapter is not only to discuss the psychophysical methods relevant to this particular study but to discuss also the general instrumentation of such studies on a large scale digital computer. This goal will be approached by first discussing the theory of linear systems and the extension of the

theory to discrete systems. Computer terminology and concepts must of necessity be used throughout the discussion. Readers unfamiliar with computers are referred to Appendix B.

The Analysis of Linear Systems

The analysis of linear systems has been discussed by Papoulis (1962), Guillemin (1963), and others. In review, a physical system may be analyzed by studying the relationship between the input to the system, $f(x)$, and the output, $g(x)$. This relationship, illustrated in Figure 1, may be expressed mathematically as

$$S[f(x)] = g(x), \tag{1}$$

where S is a transformation of $f(x)$ into $g(x)$.

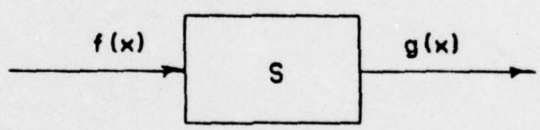


Figure 1. Representation of a physical system.

The system is completely characterized if the output for any input is known. In general, however, determining the output for any conceivable input is a formidable, if not impossible, task unless certain simplifying constraints are imposed on the system. A system is said to be _linear_ if

$$L[f_1(x) + f_2(x)] = L[f_1(x)] + L[f_2(x)] = g_1(x) + g_2(x) \tag{2}$$

and

$$L[c\ f(x)] = c\ L[f(x)] = c\ g(x) \tag{3}$$

for all inputs, $f(x)$, and constants, c. L symbolizes the linearity property. Property (2) is called superposition, which indicates that L processes the additive inputs as if they were processed separately and then added. Property (3) is called scaleability, which indicates that the outputs are scaled in correspondence with the inputs. The mathematical operations represented in (2) and (3) are vector addition and scalar multiplication.

Another simplifying property is that of stationarity. A system is stationary if

$$S[f(x - xo)] = g(x - xo), \tag{4}$$

which is to say that the output for a given input will be independent of a shift in the independent variable. The property of stationarity is more properly termed time invariance when the independent variable is time.

It can be shown that a system which has been doubly constrained to be both linear and stationary can be completely specified by a single function, $h(x)$, the impulse response of the system. The impulse response is the output of a system when its input is an impulse, $d(x)$.

$$L[d(x)] = h(x) \tag{5}$$

As approximate examples of impulse responses one may tap his cheek as the articulators are set to produce various vowels. The oral cavity forms a physical system which when struck produces a

characteristic pressure wave at the mouth corresponding to the given vowel. Striking one's desk results in the production of a sound wave which approximates its impulse response. These samples are only approximate because the systems are not excited with true impulses, which by definition are infinitely sharp.

The magnitude of the impulse response will be proportional to the strength of the impulse; the stronger the impulse, the greater the response.

A given input signal may be conceptualized as a sequence of infinitely narrow impulses whose individual power corresponds to the amplitude of the waveform at the individual points to which they temporally correspond. Each impulse sequentially applied to the input of a physical system would cause the system to respond by generating an impulse response scaled in accordance with the power of the impulse. Inasmuch as the impulse response to a given impulse may not have died away before the next impulse is applied, the output will consist of the algebraic sum of each sequentially produced and scaled response.

The process described above is called <u>convolution</u>. The input function is said to be convolved with the impulse response to produce the output and is represented mathematically as

$$g(x) = \int_{\infty}^{\infty} f(\lambda)h(x - \lambda)d\lambda \tag{6}$$

and

$$g(x) = \int_{\infty}^{\infty} h(\lambda)f(x - \lambda)d\lambda. \tag{7}$$

The equations show that the order of convolution is immaterial. It should be noted that the convolution integral applies only to linear, stationary systems.

Convolution is only one approach to the analysis of linear, stationary systems. In convolution, the input is broken up into "slices," the output determined for each slice, and the individual outputs sumed to form g(x). The summing process is difficult because scaled and temporally displaced copies of the impulse response must be algebraically added. The Fourier Integral is important in the analysis of linear, stationary systems because of the ease with which the output may be determined. As introduced, the Fourier transform converts a time domain representation of a waveform into a frequency domain representation. The Fourier Integral transform is defined mathematically as

$$F(w) = \int_{\infty}^{\infty} f(x)e^{-jwx}\,dx. \qquad (8)$$

F(w) is in general complex. The units of w are radians per unit of x. Although the Fourier transform does not exist for all functions, it will for those functions of interest to this thesis. The inverse Fourier Integral transform is given by

$$f(x) = 1/2\pi \int_{\infty}^{\infty} F(w)e^{jwx}\,dw. \qquad (9)$$

To appreciate the application of the Fourier integral to system analysis, one must realize that if the input to a system is an exponential e^{jwx}, then the output will also be an exponential,

but scaled in proportion to the input.

$$L[e^{j\omega x}] = ke^{j\omega x} = g(x) \tag{10}$$

The proportionality constant, k, is in general complex and is usually symbolized $H(\omega)$. The proportionality constant is represented in both polar and rectangular coordinates in (11).

$$H(\omega) = A(\omega)e^{j\theta(\omega)} = R(\omega) + jX(\omega) \tag{11}$$

$H(\omega)$ is the Fourier transform of the impulse response, $h(x)$, and is called the <u>system function</u>.

$$H(\omega) = \int_{\infty}^{\infty} h(x)e^{-j\omega x}\,dx \tag{12}$$

With $F(\omega)$ the Fourier transform of $f(x)$ and $G(\omega)$ the Fourier transform of $g(x)$ it can be shown that

$$G(\omega) = F(\omega)\,H(\omega), \tag{13}$$

which is to say that in the frequency domain the output of a linear system is the product of the input and the system function. The value of the Fourier transform in analysis can be seen from (13); the Fourier transform maps convolution into multiplication. The added cost of such analysis is two forward transforms and one inverse transform.

The Discrete Fourier Transform and Discrete Linear Systems[1]

Computers are only able to process numbers or symbols because of their inherently discrete nature. Symbolic processing is of little use in signal processing, however, because signals of interest generally have no known representative formull. It is consequently necessary to represent continuous signals as though they were discrete. Discrete functions may be represented within computer memory, inasmuch as such functions are defined only at discrete values of the independent variable, by simply storing sequential values of the function in contiguous memory locations. Shannon has bridged the gap between the representation of continuous function as discrete function by showing that continuous functions may be sampled, i.e., measured at discrete intervals of time, without loss of information, provided that the sampling rate is greater than twice the frequency of the highest frequency component within the function.

In addition to discretely representing signals of interest, it is also necessary to develop discrete forms of the Fourier transforms such that the properties of the transforms may be applicable to digital signal processing. The Fourier transform of a discrete function, called the discrete Fourier transform (DFT) must not only be a discrete function itself, but also be bounded such

[1]It is difficult to exercise much originality in a concise discussion of discrete linear systems due to the ever increasing number of references on the subject. This particular introduction to discrete systems follows partially from the work of Cole (1973).

that it may be represented in the computer.

It is instructive to note in regard to the development of the DFT that it may be shown that the transform of a periodic waveform is discrete whereas the transform of an aperiodic waveform is continuous. Inasmuch as the DFT of a discrete function must also be discrete, it logically follows that the function and its transform must also be periodic. Cole (1973) has noted that many of the problems encountered in digital signal processing are due to the failure to remember the periodicity of the DFT and its inverse.

Rabiner and Schafer (1969), in an approach to the development of the DFT, first developed the so called Z-transform of a discrete, aperiodic function. By the properties just introduced, the Z-transform must be both periodic and continuous. The Z-transform is sampled to produce a periodic, discrete function, which is taken to be the DFT.

Let f_J be a discrete periodic function of period N. The DFT of f_J is another discrete periodic function, F_K, also with period N.

$$F_K = \sum_{J=0}^{N-1} f_J \, e^{-j(2\pi/N)JK} \qquad (14)$$

The complex expression $e^{-j(2\pi/N)}$ is the principal Nth root of unity and is often abbreviated as W.

$$F_K = \sum_{J=0}^{N-1} f_J \, W^{-JK} \qquad (15)$$

The inverse DFT is similarly defined.

$$f_J = 1/N \sum_{K=0}^{N-1} F_K W^{JK} \tag{16}$$

Signal processing became practical in 1965 with the development of the Fast Fourier transform (FFT) by Cooley and Tukey. The FFT is a very efficient algorithm for computing the DFT and as such does not represent a different type of transform. The computation time required to compute the DFT of an N sample function is proportional to N squared. The FFT is able to perform the same transform by eliminating redundant calculations through matrix factorization in a time proportional to N log N, a speed up of N / log N.

A discrete linear system is one which has discrete functions as its input, output, and impulse response. Discrete linear systems also have the properties of superposition and scaleability. If the system is also stationary, its input and output will be related by discrete aperiodic convolution, which is defined mathematically as

$$g_J = \sum_{K=-\infty}^{\infty} f_K h_{J-K} . \tag{17}$$

A second kind of discrete convolution is periodic, or circular convolution. The circular convolution of two periodic discrete functions, f and h, is a third periodic function, g,

expressed mathematically as

$$g_J = \sum_{K=0}^{N-1} f_K h_{J-K} . \qquad (18)$$

It should be noted that f, h, and g all have the same period, N.

The circular convolution is related to the DFT in that it can be shown that if F, H, and G are the respective discrete Fourier transforms of f, h, and g that

$$G_K = F_K H_K . \qquad (19)$$

Most real world signals are generally aperiodic. Consequently, in order to use the DFT for practical signal processing, it is necessary to implement an aperiodic convolution with a periodic convolution. Stockham (1966, 1969) has discussed such a method. The process basically involves appending enough zeros to f and h so as to give them the same period and to ensure that the resulting period is at least greater than one less than the sum of the original number of samples in f and h. Since the length of the aperiodic convolution of f and h is one less than the sum of the number of samples in each, the results of the periodic convolution will be periodic but will represent the aperiodic convolution of f and h.

Stimulus Recording, Storage, and Playback

Digital techniques were used to record the speech sounds used in this study. The sounds were spoken by a male speaker, 33 years of age, who was judged by the author to have good voice quality, into a one-inch B and K condenser microphone. The electrical signal was amplified, low-pass filtered at 4 kHz, and sampled at 10,000 samples per second with 14-bit resolution. The samples were stored packed two samples per word on disk.

The disk, as a mass storage device, only allows data to be addressed, i.e., referenced, in blocks of 128 words. The processed signals used as stimuli in the study, however, are not necessarily integral number of blocks in duration. Development of a disk audio file system incidental to the study was consequently not only expedient but necessary because of the required control of signal onset and duration, and because of the need to reference quickly a large number of signals throughout the study.

The audio file system facilitated the storage and retrieval of audio data on disk by automating the bookkeeping associated with storage and retrieval such as the location and duration of the data on disk, the sampling frequency, etc. The system was implemented by storing a directory image on a reserved portion of the disk. The directory image contains slots for 1023 file headers, which in turn, contain slots for the sampling frequency, beginning block location, file length, and date of file creation. A particular file is referenced by a number between 1 and 1023, which corresponds to the file header slot in the directory image. Rather than request X

words out of Y blocks beginning at block Z, the file system allows one simply to request the data by specifying the file number. The system automatically looks up "free space" on disk whenever a new file is created so as to allow one to store data simply by specifying the file number under which the data is to be stored.

The following example is given to demonstrate the file system in use and to discuss signal playback. Suppose one wished to replay signal X. The signal would be specified in the user's program. The file system would read the directory image from disk and look up the associated bookkeeping data from the appropriate slot in the directory image. Enough blocks of audio data would then be read from disk into contiguous memory locations to include the entire audio signal. The clock in the digital-to-analog converter would be set from other data in the file header. The computer would then be instructed to output the correct number of samples from the word addressable memory to the digital-to-analog converter.

The signals from the digital-to-analog converter are low-pass filtered at 4 kHz and amplified for presentation through high quality headphones.

Filters and Stimulus Preparation

Filters are linear systems in which the amplitude $A(\omega)$ and the phase $\theta(\omega)$ are both functions of frequency. The system function of a filter may be represented mathematically as

$$H(\omega) = A(\omega)e^{-j\theta(\omega)} \tag{20}$$

where $\theta(\omega)$ is known as the phase shift or phase lag function, and is defined as $-\emptyset(\omega)$ where $\emptyset(\omega)$ is the phase of the system.

The stimuli peculiar to this study are prepared by processing standard stimuli with a filter whose phase characteristic is that of the desired phase modification. It is desired that only the phase be modified so the amplitude characteristic of the filter is set to unity. Such filters are called all-pass filters and may be represented as illustrated in Figure 2.

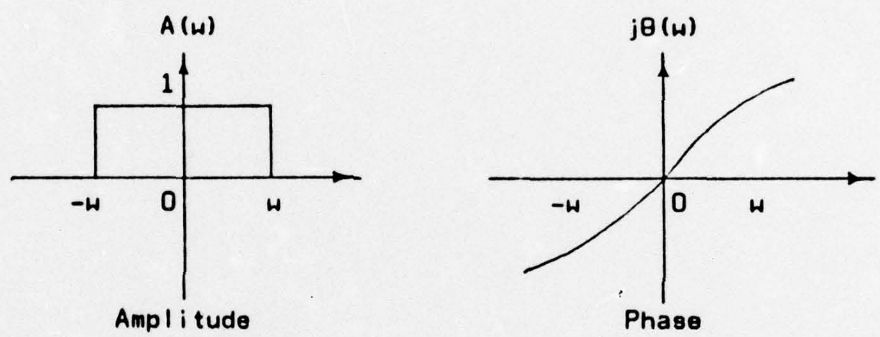


Figure 2. Amplitude and phase characteristics of a generalized all-pass system.

A systematic investigation of phase sensitivity depends on a systematic variation of the phase function. The phase modification of choice in this study is to systematically delay various frequency bands within the stimuli. The parameters of the phase modification include center frequency (FC), bandwidth (BW), and time delay (TD). The idealized delay characteristic of the system used to achieve the phase modification is illustrated in Figure 3.

It may not immediately be evident that the type of phase modification shown in Figure 3 can generally represent phase distortion. It will be shown in Chapter V, however, that any phase

distortion will cause frequency component delay. The systematic variation of a phase function incidental to investigating phase sensitivity may be more easily conceptualized in reference to the parameters of the delay. The idealized parameters of center frequency, bandwidth, and time delay may be utilized for systematic modification.

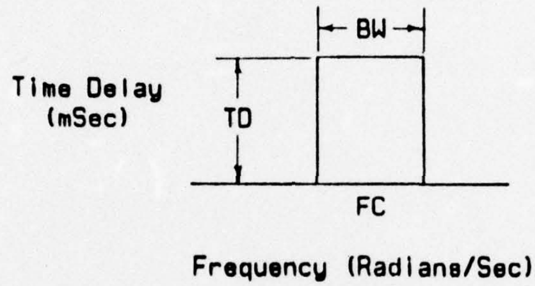


Figure 3. Delay characteristic of the system used in phase modification of signals to be investigated, showing center frequency (FC), bandwidth (BW), and time delay (TD) as parameters.

Usually in the study of phase, a discussion of the delay property follows from the analysis of phase distortion as will be done in Chapter V. It is more convenient to the development of the present discussion, however, to introduce delay before formally discussing phase distortion.

The stimuli used in this study were prepared by convolving, i.e., filtering, selected source stimuli with appropriate filters designed to introduce the required distortion. A digitally implemented high-speed convolution algorithm (Stockham, 1969) was used which was based on multiplication of the DFT's of the stimuli and filter as shown by (13).

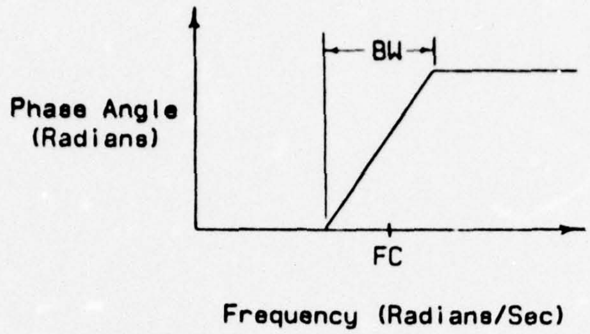


Figure 4. Idealized phase characteristic of the system to be investigated.

Filter Design

The all-pass, phase modifying filters used in this study were designed by the so called Five T's method. The filters are of the non-recursive type which is to say that they are digitally implemented to function in the frequency domain. Using the Five T's method, a desired phase characteristic is first specified in the frequency domain. It should be remembered that the specified phase characteristic is discrete but it is convenient to represent the characteristic graphically as if it were continuous, as illustrated in Figure 5. The realized phase is continuous.

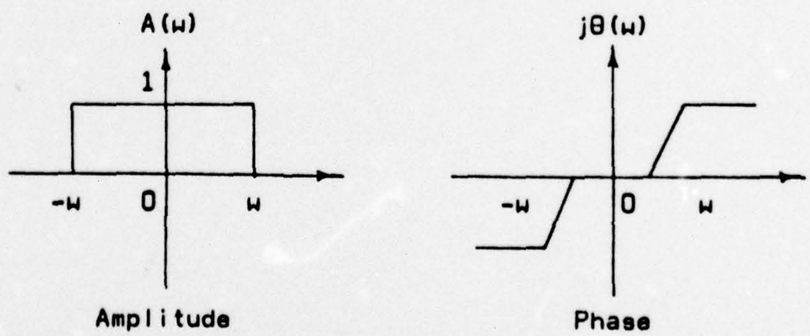


Figure 5. Frequency domain representation of the idealized all-pass system to be designed.

It is natural to specify the desired phase characteristic in polar coordinates inasmuch as it is in polar coordinates that the effects of the characteristic are conceptualized. The DFT algorithm, however, requires that the data be represented in rectangular coordinates so the next step is to convert to the rectangular coordinate system. The data is next transformed (T1) to the time domain by the inverse DFT to produce the impulse response of the filter. The time domain representation is guaranteed to be real if the magnitude is specified to be even and the phase to be odd as illustrated in Figure 5.

Another peculiarity of digital signal processing is that the impulse response of the desired filter may be too long to be contained in the computer's memory. Inasmuch as the memory size of the computer is finite, the impulse response must be truncated (T2). To reduce the traumatic effects of an abrupt truncation, the impulse response is windowed, i.e., taylored (T3), so as to turn it on and off smoothly. The modified impulse response is saved tentatively on disk for possible later use in signal processing the various test stimuli. The impulse response is transformed (T4) back to the frequency domain by the inverse discrete Fourier transform, converted to polar coordinates, and then compared, i.e., tested (T5), against the originally specified characteristic. The impulse response is retained for later use in processing the test stimuli if the characteristic of the approximated filter is close enough to the specified ideal filter.

Digital and Analog Equipment

All digital and analog equipment required for the study was available to the author through the Sensory Information Processing Group (SIPG), with which the author is associated at the University of Utah. The main computer facility includes a single-user PDP-10 computer with 64 K of 36-bit words, a PDP-10 time-sharing computer with 196 K of words for program preparation and debugging, and all associated computer peripherals. All audio equipment, including filters, amplifiers, tape recorders, etc., is of the highest professional quality. Included in the SIPG facilities is a large sound isolated "quiet room" in which test signals were recorded and the psychophysical listening tests were conducted.

Manipulation of the audio signals was facilitated by use of a general purpose audio console designed and fabricated by the SIPG staff engineer (Warnock, 1973). The audio system consists of a modular set of high quality amplifiers, filters, attenuators, and signal generators that may be plugged into the console cabinet. The cabinet contains a common power supply. Signal routing is handled through BNC connectors on the front of each module. Various grounding configurations among the modules, cabinet chassis, and earth-ground are provided by means of switches so as to minimize hum and noise within the system.

The variable gain amplifier modules allow continuous gain adjustments from -40 to +20 dB over a frequency range from DC to 50 kHz.

The selectable low-pass filter module allows switch selection of one of four TT Electronics (Model J77A) low-pass filters with 3 dB cutoff frequencies of 4 kHz, 7 kHz, 10 kHz, and 15 kHz. The attenuator module is patterned after the HP model 350D and allows attenuation up to 110 dB in 1 dB steps. The digital-to-analog converter has 16-bit resolution with an output voltage range of 20 Volts. The output of the converter is attenuated 17 dB to avoid overloading the low-pass filter, filtered at 4000 Hertz, and then amplified to drive the headphones. The low-pass filter introduces a loss of 6 dB. No calibration data was available for the Koss PRO-4A headphones used throughout the study.

The signal-to-noise ratio of the amplifiers and filters is greater than 100 dB. The total harmonic distortion of the amplifiers is less than 86 dB. The arrangement of the equipment is illustrated in Figure 6.

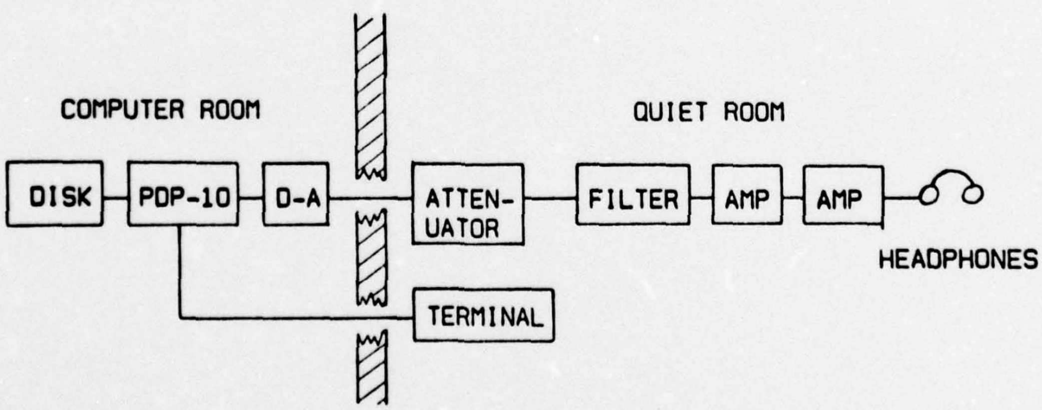


Figure 6. Equipment arrangement for discrimination testing of dispersion.

Psychophysical Methods

Psychophysical testing of phase is difficult because of the nature of the required discrimination task, which is the detection of changes in sound quality. Various factors influence the cognitive process of assigning a metric to the quality of an experience, some of which may be quite independent of the presence or intensity of the various physical dimensions of the event under consideration.

Listeners generally have little difficulty in assigning values to single dimensional events such as to sound frequency or intensity. The discriminability of sound quality, however, is more difficult because of the multi-dimensionality of the factors that influence the psychological evaluation of the event. The detection problem may be further complicated by the possibility that different aural mechanisms account for different aspects of phase sensation.

The psychophysical method used in determining the sensitivity of the ear in this study was that of <u>constant stimuli</u> in which a direct comparison was required between a standard stimulus and a number of phase modified variants of the stimulus. Notwithstanding the difficulty of establishing a criterion for phase changes, listeners were instructed to make phase change discriminations in the series of experiments to be discussed by responding to "any difference in the way two stimuli sounded" when presented together as a stimulus pair.

A signal comparison testing program was written by the author that randomly presents a number of audio file pairings to a

listener for discrimination. The audio file numbers of the files to be paired are specified in a command list so as to make the testing program as general as possible. For example, if audio file 103 and 413 are to be compared with file 87, the command list would contain the following data:

87, 103
87, 413
...

A random-like number generator is used within the testing program to randomize the command list and consequently the order of the signal pair presentation. The number of times the testing program randomizes and executes the command list may be set by a program parameter. The temporal spacing between the signals of the pair is also adjustable.

The signal pairs were presented to each listener during the course of a test through high quality headphones at "average conversational speech levels." With each presentation of a signal pair, the listener was required to respond either by depressing a "same" or "different" key on the computer terminal depending on whether or not the signals sounded the same or different. The testing program records the listeners responses and provides a statistical reduction of the data at the conclusion of the test.

CHAPTER V

THE ANALYSIS OF PHASE DISTORTION

It will be helpful to review the relation between phase characteristics and phase distortion as obtained by analytical methods before discussing experimental results. Phase modifications are secured by processing a signal, f(t), with a phase modifying system. The absolute phase characteristic of f(t) is of no interest except to the extent that it joins in the definition of f(t). The phase characteristic of interest is that of the phase modifying system which is represented in the modified signal, g(t), as the phase difference between g(t) and f(t). It is consequently convenient to discuss the phase characteristic of the modifying system, H(w), as the phase characteristic of interest.

Characteristics of a Distortionless System

Following the development of Papoulis (1962), a distortionless filter is one whose output, g(t), to an arbitrary input, f(t), has the same form as the input. Noting that scaling or delaying a signal does not change its form, the input-output relationship of a distortionless filter may be written

$$g(t) = Af(t - t_o). \qquad (21)$$

The system may be analyzed using the Fourier integral transform. With G(w) the transform of g(t) and $e^{-j\omega t_o} F(\omega)$ the transform of

f(t - to), then

$$G(\omega) = Ae^{-j\omega to} F(\omega). \tag{22}$$

Following from (13), the system function of a distortionless system is given by

$$H(\omega) = Ae^{-j\omega to}, \tag{23}$$

from which the amplitude is seen to be constant and the phase linear.

$$A(\omega) = A \qquad \theta(\omega) = -j\omega to \tag{24}$$

The system is said to be amplitude distorted if A(ω) is not constant and phase distorted if θ(ω) is not linear.

It may be observed from (24) that the time, t, by which the system delays the input is given by

$$to = \theta(\omega)/\omega = d/d\omega\, \theta(\omega), \tag{25}$$

which is to say that distortionless systems act only to delay the input signal by an amount equivalent to the slope of the phase characteristic. If θ(ω) is in radians and ω = 2 f, where f is frequency in Hertz, the delay will be in seconds.

Interpretation of Phase Distortion

Phase distortion occurs when a system fails to maintain or shifts the phase relations among the frequency components of an applied signal. Consider first the effect of adding a constant

phase, θ, to the phase characteristic. Letting $A(\omega) = 1$, a distortionless filter is represented by

$$H(\omega) = e^{-j\omega t_o} \qquad (26)$$

as developed in the previous section. Adding the constant phase, θ, gives

$$\hat{H}(\omega) = e^{j(\theta-\omega t_o)} = e^{j\theta} e^{-j\omega t_o} = e^{j\theta} H(\omega), \qquad (27)$$

which is to say that shifting the phase characteristic by a constant phase is equivalent to multiplying the system function by a complex constant. Lane (1930) and Steinberg (1930) have carried the interpretation further. Since

$$e^{j\theta} = \cos\theta + j\sin\theta, \qquad (28)$$

and noting that $\theta(\omega)$ is an odd function, the phase shifted output may be considered to be the sum of two parts. The first part is a fidelity copy of the input signal, $f(t)$, scaled in amplitude by $\cos\theta$ and delayed by t_o. The second part results from shifting all the components of $f(t)$ by $\pi/2$, scaling by $\sin\theta$, and then delaying by t_o.

Two special cases should be noted. If θ is 0 or an even multiple of π then $e^{j\theta} = 1$, which is equivalent to a distortionless system. If θ is an odd multiple of π then $e^{j\theta} = -1$, which is equivalent to a distortionless system with reversed output. In either special case there is no distortion, only a delay.

In conclusion, the slope and y-intercept components of a linear phase function correspond respectively to a distortionless delay and a distortion producing constant phase addition resulting from the constant phase addition. It is convenient to conceptualize the two components as causing sequential operations in linear systems, i.e., a delay followed by a distortion.

Distortion Following from Nonlinear

Phase Characteristics

Steinberg (1930) analyzed the effect of phase distortion resulting from curved phase characteristics by considering such characteristics to be limiting cases of characteristics made up of a number of straight lines. Each line approximates the curved characteristic for a given frequency range, $\Delta\omega$. The frequency components in each range will be subjected to a delay and distortion as previously discussed and illustrated in Figure 7.

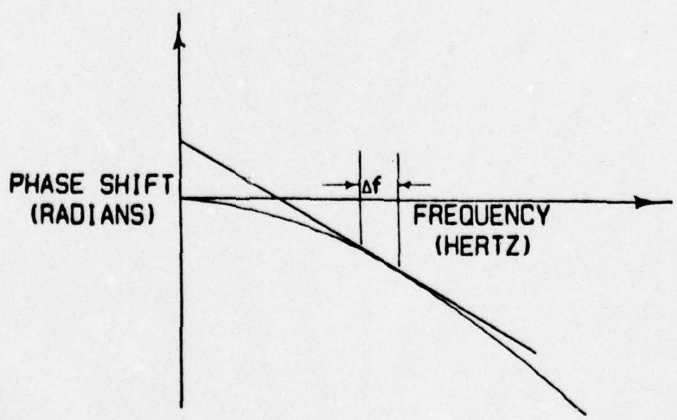


Figure 7. Curved phase characteristic segment approximated by a straight line. The slope of the line represents the temporal delay of the frequency band and the Y-intercept the constant delay.

It should be apparent when considering the effect of a curved phase characteristic on the whole input signal that each band is delayed differently than its adjacent bands so as to spread out the signal on the time scale. The signal is said to be dispersed. Each band of frequencies, $\Delta\omega$, is delayed relative to the minimum slope of the phase characteristic. The relative delay has been defined by Steinberg as delay distortion.

CHAPTER VI

SENSITIVITY TO DISPERSION IN IMPULSES

The purpose of the experiments described and discussed in this chapter was to determine the discriminability between an impulse and impulses in which a band of frequency components are delayed. An impulse is an infinitesimally short signal in which all frequency components are present at equal amplitudes. Impulses were chosen as interesting stimuli because they appear to be among the simplest stimuli possible inasmuch as their spectra are of constant amplitude. It is not unreasonable to assume that the psychophysical data derived from tests using more simple stimuli may be more easily analyzed than data derived from those tests using more complex stimuli such as speech. Such data may lead to more direct insights into the nature of phase discriminability. The insights derived using simple stimuli may aid in the analysis of data derived from tests utilizing the more complex stimuli.

Sensitivity of Three Subjects to Dispersion in Impulses

The intent of the first experiment was to ascertain the discriminability of three subjects between an impulse and impulses in which a band of frequency components are delayed. The manner in which a given impulse used as a stimulus in the experiment was modified has been discussed in Chapter IV. In brief, a standard

impulse is processed so as to delay a band of its constituent frequency components, the independent variables of interest being center frequency, bandwidth and temporal extent of the delay. The center frequency in the present experiment, however, was held constant at 500 Hertz. Four values of bandwidth and four of delay were selected so as to test discriminability as a function of 16 unique combinations of bandwidth and delay. The bandwidths included 100, 200, 400 and 800 Hertz. The delays included 0.0625, 0.125, 0.25 and 0.5 milliseconds. Sixteen phase modified impulses were prepared according to the prescribed schedule as stimuli for the experiment.

The experiment consisted of administering 16 individual tests to each of the three subjects so as to obtain a phase discriminability estimate for each of the scheduled phase conditions. Each test consisted of a sequential presentation of stimulus pairs. The subjects were required to make a forced choice decision as to whether or not the stimuli of a given test pair were the "same" or "different." Either stimulus of the pair could be the standard impulse (S1) or a phase modified variant of the impulse (S2) so as to generate four unique stimulus pair types, i.e., S1-S1, S1-S2, S2-S1 and S2-S2. The probability of occurrence of each type was equal so as to represent a rectangular distribution. This scheme was adopted to average out any experimental bias. One-hundred stimulus pairs were presented in each test.

Three female speech pathology students with normal hearing and of 21 to 23 years of age served as subjects. The stimuli were

presented monaurally to the subject's left ear at a "low" sensation level of approximately 30 to 40 dB. The intra-stimulus duration between stimuli of the pair was approximately 700 milliseconds. The inter-stimulus duration was 500 milliseconds, i.e., 500 milliseconds elapsed after the subject's response and the automatic onset of the next stimulus pair.

The subjects were sequentially tested at the same phase condition for each of the 16 tests. The test order was randomized. This approach was taken because it was not initially certain that the range of values selected for the independent variables would be of interest for all subjects. The approach allowed modification of the variable values without detriment to the experiment. The loss sustained by the experiment because of this approach, however, was the loss of absolutely identical testing conditions for each of the tests inasmuch as the formal testing period required three testing sessions over as many days. The subjects received three training session prior to formal testing.

The test results for each of the three subjects are tabulated in Table 2 (Appendix C) and are illustrated in Figures 8 and 9. The results of the study are interesting but not overly surprising. The data when plotted as a function of delay as shown in Figure 8 indicate a definite dependence of discriminability on delay for a given bandwidth. It is not clear from a consideration of the data plotted as a function of bandwidth as shown in Figure 9 that there is a significant increase in discriminability with bandwidth for a given delay.

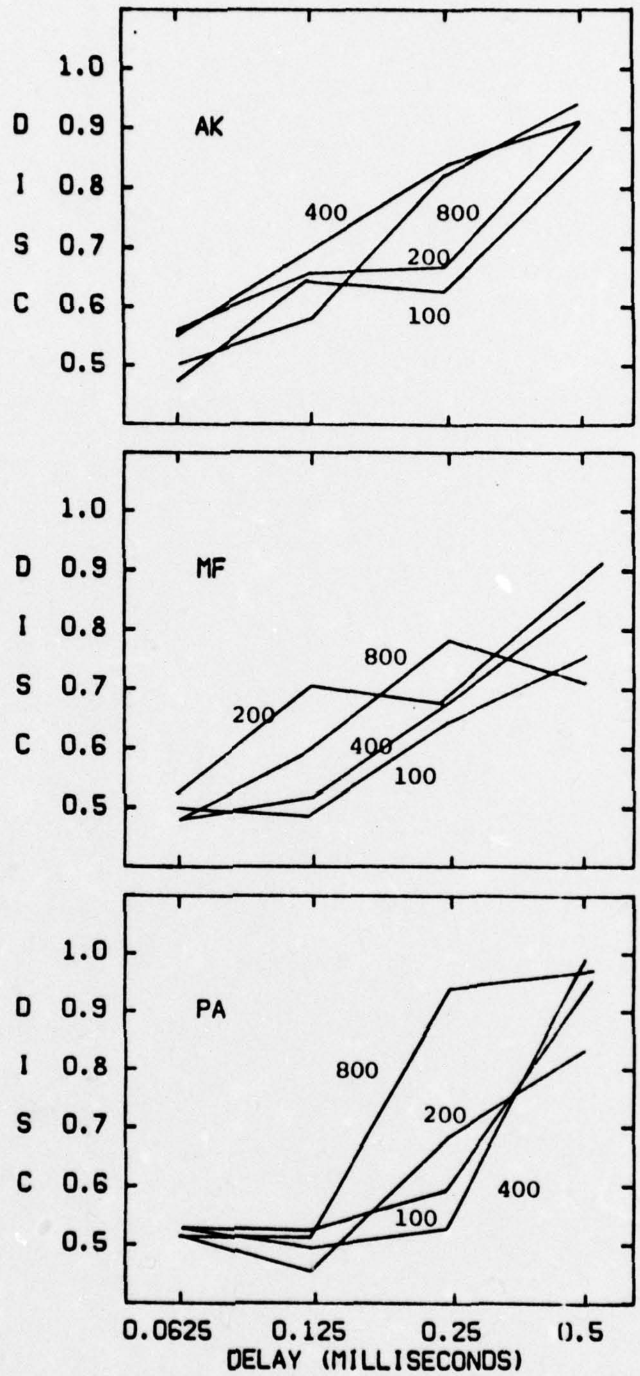


Figure 8. Discriminability of subjects AK, MF and PA between an impulse and phase modified impulses plotted as a function of delay with bandwidth as parameter. Center frequency of the delayed band is 500 Hertz.

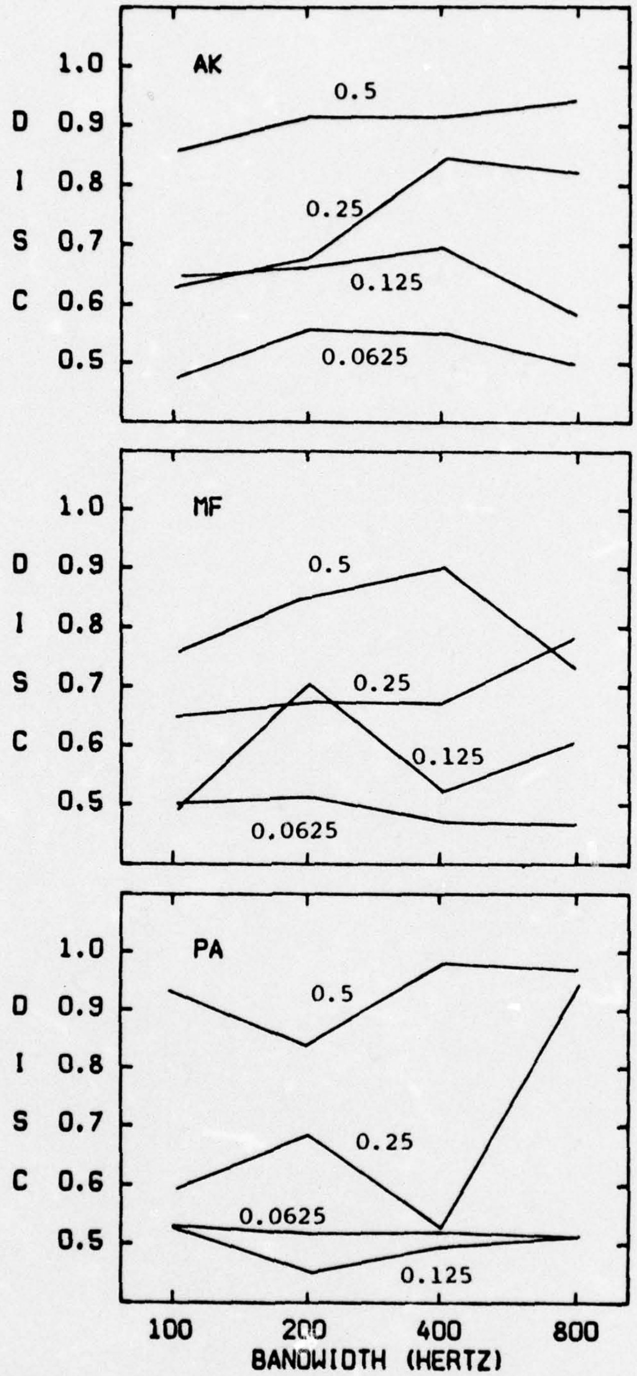


Figure 9. Discriminability of subjects AK, MF and PA between an impulse and phase modified impulses plotted as a function of bandwidth with delay as parameter. Center frequency of the delayed band is 500 Hertz.

The only similarity among the subjects is an increase in discriminability with bandwidth at a delay of 0.25 milliseconds. The curves appear relatively flat otherwise except for a dip in the 0.5 millisecond curve for subject MF and PA at different bandwidths, and a peak in the 0.125 curve for subject MF.

A convenient statistic for indicating the ability to discriminate is half way between chance and perfect performance, i.e., the 0.75 discrimination score. The subjects are able, in general, to respond to delays in impulses of between 0.125 and 0.5 milliseconds which are, apparently, relatively independent of bandwidths greater than 100 Hertz.

Sensitivity to Dispersion in Impulses as a Function of Intensity

The test stimuli used in the experiment just discussed were presented at a low level to prevent possible overloading and subsequent distortion in the headphones. It is not impossible that the stimulus level was so low that the stimuli could not be heard adequately to provide valid discriminations.

A second experiment was conducted to test the effect of intensity on discriminability. A standard impulse was paired with a phase modified impulse of 500 Hertz center frequency, 400 Hertz bandwidth and 0.25 millisecond delay at three intensity levels. The intensity levels were 6 dB below the level used in the last experiment, at the level used in the last experiment, and 6 dB above the level used in the last experiment. The testing design and conditions were otherwise identical to the experiment just

discussed. The test results are tabulated in Table 3 (Appendix C) and are illustrated in Figure 10. The discrimination scores may be observed to be definitely related to intensity.

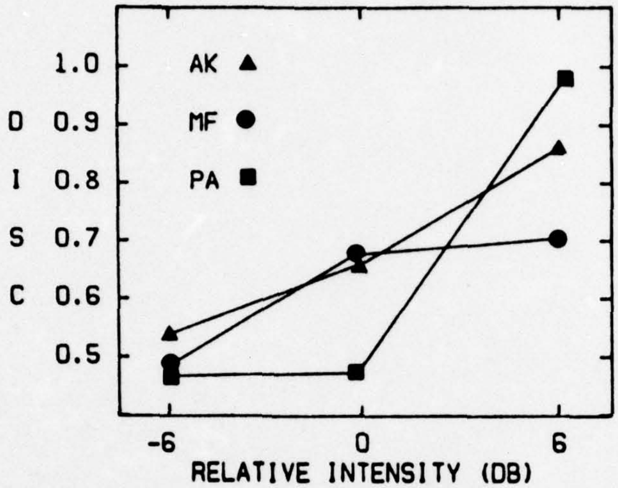


Figure 10. Discriminability of subjects AK, MF and PA between an impulse and phase modified impulses plotted as a function of intensity. The parameters are 500 Hertz center frequency, 400 Hertz bandwidth and 0.25 milliseconds delay.

To test discriminability at higher sensation levels and to obtain a more complete set of data in which discriminability is determined as a function of intensity, the author repeated the experiment serving as his own subject. Sixty stimulus pairs were used in each test. The intra-stimulus interval was set to approximately 800 milliseconds and the inter-stimulus interval to 500 milliseconds. The stimuli were presented monaurally to the right ear. The audio equipment was set to provide approximately 5 dB gain following the output of the digital-to-analog converter, i.e., 17 dB attenuation before filtering and 22 dB gain following filtering. An attenuator was included in the circuit to provide for intensity level control. Using the method of adjustment, the author

determined his threshold to the standard impulse to be at 93 dB attenuation. It is in reference to this setting of the attenuator that the various sensation levels used in the experiment are referenced. The experimental results are tabulated in Table 4 (Appendix C) and are illustrated in Figures 11 and 12.

Discriminability may be observed from Figures 11 and 12 to be strongly dependent on intensity for a given center frequency, bandwidth and delay. This general finding is in agreement with the previous experiment in which three subjects were tested. The widely separated parametric curves of Figure 11 indicate a strong dependence on delay for a given bandwidth whereas the close clustering of the parametric curves in Figure 12 indicate little, if any, dependence on bandwidth for bandwidths greater than 100 Hertz for a given intensity. The intensity function illustrated in the different sub-figures of Figure 11 appear to be roughly linear with slopes that appear to have a slight dependence on delay. One would expect increased discriminability as a function of intensity inasmuch as the dispersed components of the impulse become more and more audible for a given dispersion with an increase in signal level. Greater delays in the dispersed frequency band have already been shown to be more discriminable than shorter delays. It follows, consequently, that the slope of the intensity function should be dependent on the given delay. That the stimuli did not produce nonlinearities is implied by the well behaved nature of the data. Speech stimuli presented at the same levels did not have the buzzy quality characteristic of distortion overloading.

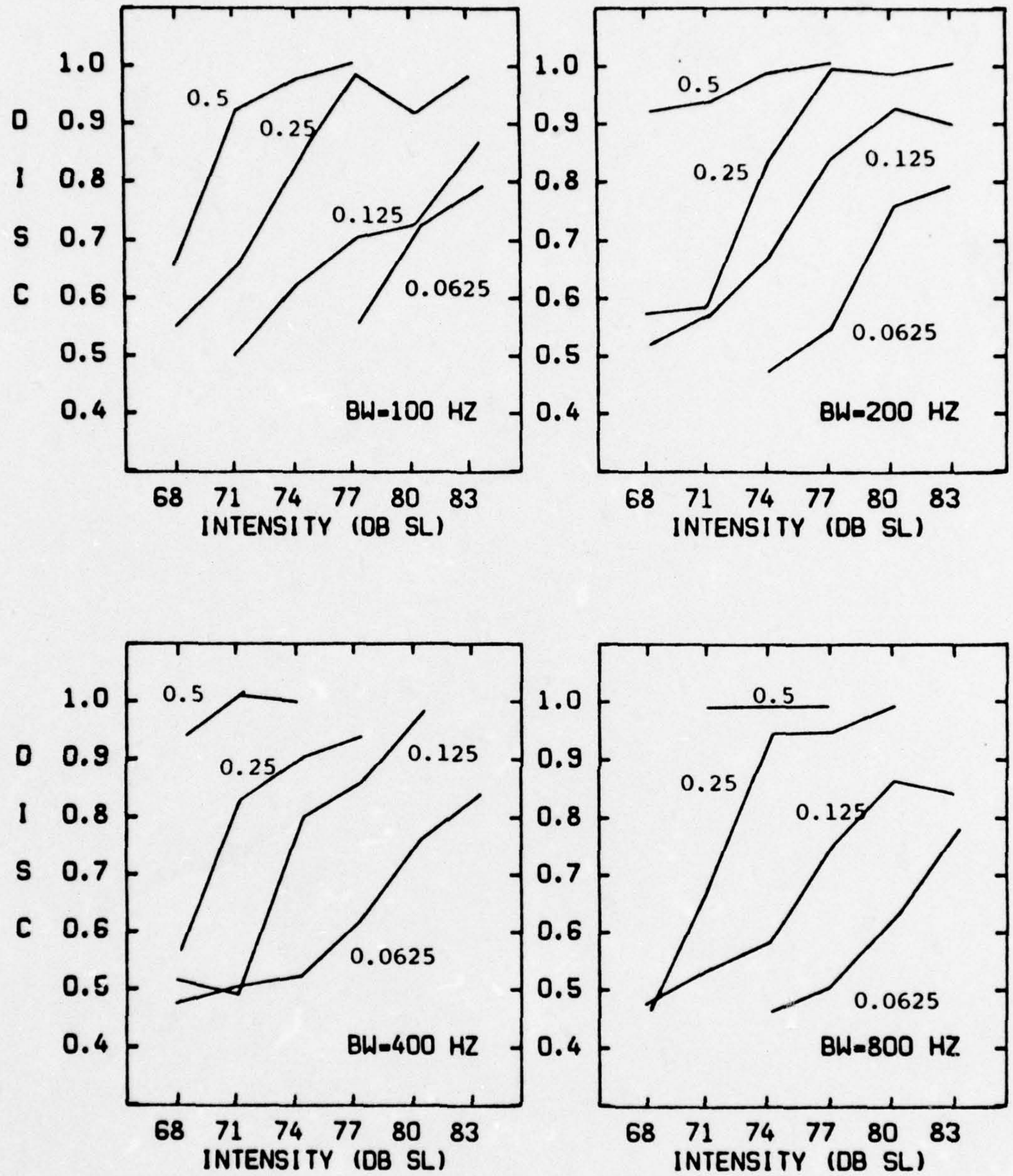


Figure 11. Discriminability of Subject WG between an impulse and phase modified impulses as a function of intensity with delay as parameter.

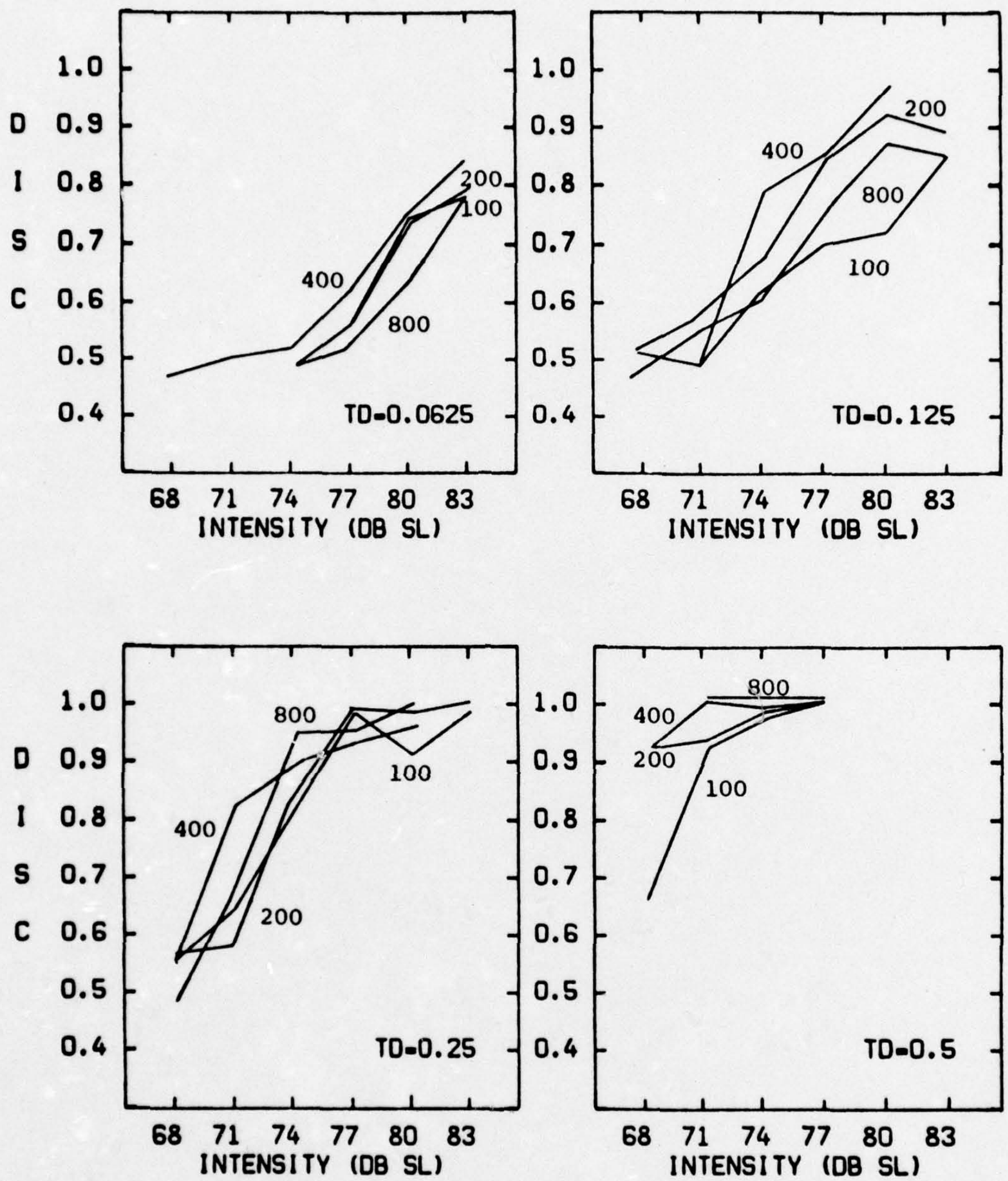


Figure 12. Discriminability of Subject WG between an impulse and phase modified impulses as a function of intensity with bandwidth as parameter.

Discriminability as a Function of the Psychophysical Method

It is well known that test results are a function of the testing paradigm. The experimental results derived thus far have been derived from separate tests for each data point as already discussed. A second test series in which more than one stimuli was randomly presented during the course of the tests was conducted by the author with himself as subject to obtain an estimate of how the results might be affected by the testing method.

Eight separate tests were administered to reduce the length of any given testing session. Each test tested four of the possible 16 phase conditions. The four phase conditions were randomized across all the phase conditions and enough tests were administered to obtain two estimates for each of the phase conditions. The test conditions were the same as the last test except for the testing method, the use of 40 stimulus pairs per condition, and the presentation of the stimuli at a constant level of 80 dB sensation level. The average of the two estimates for each condition are tabulated in Table 5 (Appendix C).

One may observe by comparing the results tabulated in Table 4 and 5 that use of the second test design results in scores indicative of less sensitivity to phase changes than that indicated by the first test design. This finding is not unreasonable. The subject is better able to concentrate on smaller nuances of difference between the standard stimulus and the modified stimulus when only one phase condition is being contrasted in a given test.

The results of the two testing designs do not appear to differ significantly.

Sensitivity of the Author to Dispersion in Impulses

To obtain a more complete set of data in which discriminability was to be determined as a function of center frequency as well as bandwidth and delay, the author repeated the first experiment discussed in this chapter but served as his own subject. A second reason for repeating the experiment was to use stimuli at higher intensities similar to those encountered in conversational level speech. A third reason for repeating the experiment was to obtain some measure of variability of the discrimination function.

The experimental design was the same as in the first experiment except that the stimuli were presented monaurally to the right ear at a sensation level of 80 dB. Fourty stimulus pairs were presented in each test. The intra-stimulus delay was approximately 800 milliseconds. The vast majority of the individual phase conditions were randomly tested three times during the course of the experiment to obtain an estimate of the variance of the discriminability function for the particular phase condition. In some cases where discrimination was obviously perfect, only two tests were performed to obtain the estimate. As many as five tests were used where variability appeared to be excessive. All the individual tests for the various phase conditions were randomly administered. The test results are tabulated in Table 6 (Appendix

C) and are illustrated in Figures 13 through 18. The intent of the author was to discuss the data trends as they were made manifest by plotting the data as functions of the three independent variables and then to discuss the trends.

Figure 13 and 14 represent the data plotted as a function of delay. Both figures indicate that discrimination is strongly dependent on delay for a given center frequency and bandwidth. The clustering of the parametric curves in Figure 13 indicates a weaker, if any, dependence of discriminability on bandwidth for the bandwidths tested. Note that in the sub-figures of Figure 13 that except for the 50 and 100 Hertz parametric curves that the curves monotonically increase to the upper, left of the sub-figures. The general clustering and monotonic increase indicate a mild but significant dependence of discrimination on bandwidth. The significance of the exceptional behavior of the 50 Hertz curve will be discussed in the next section. The locus of the parametric cluster for the 250 Hertz sub-figure, being different from that of the other sub-figures, indicates a dependence of discrimination on center frequency. This dependence on center frequency is made more evident in Figure 14 where the data is plotted with center frequency as parameter. In every case where the 250 Hertz center frequency curve is parametrically plotted, the curve is seen to be more or less isolated from the other parametric curves. The clustering of the other curves indicate a stronger dependence of discriminability on center frequency for lower frequencies that becomes less significant with an increase in frequency. The general monotonic

increase in the values of the parametric curves in all the sub-figures of Figure 14 indicate that center frequency does not become insignificant, however, at the highest frequencies.

Figure 15 and 16 represent the data plotted as a function of bandwidth. In general, the curves of both figures are slightly monotonically increasing with bandwidth. The wide but monotonic dispersion of the parametric curves in Figure 15 and the lack of similarity of the sub-figures in Figure 16 indicates a strong dependence of discrimination on delay. The 0.0625 millisecond parametric curve in all the sub-figures of Figure 15 except for the 250 Hertz sub-figure appear to be more isolated from the other parametric curves which tend to be more clustered.

Figure 17 and 18 represent the data plotted as a function of center frequency. Both figures imply a strong dependence of discriminability on center frequency for a given bandwidth and delay. The lack of general similarity of the sub-figures in Figure 17 and the dispersion of the parametric curves in each sub-figure of Figure 18 again indicates a dependence of discrimination on delay. The monotonic but moderate clustering of curves in each sub-figure of Figure 17 and general similarity of the sub-figures of Figure 18 indicate a mild but significant dependence of discrimination on bandwidth. The 0.0625 millisecond parametric curve in all the sub-figures of Figure 18 is generally isolated from the other parametric curves which tend to be more clustered except in the 100 Hertz sub-figure. This isolation is similar to that shown in Figure 15.

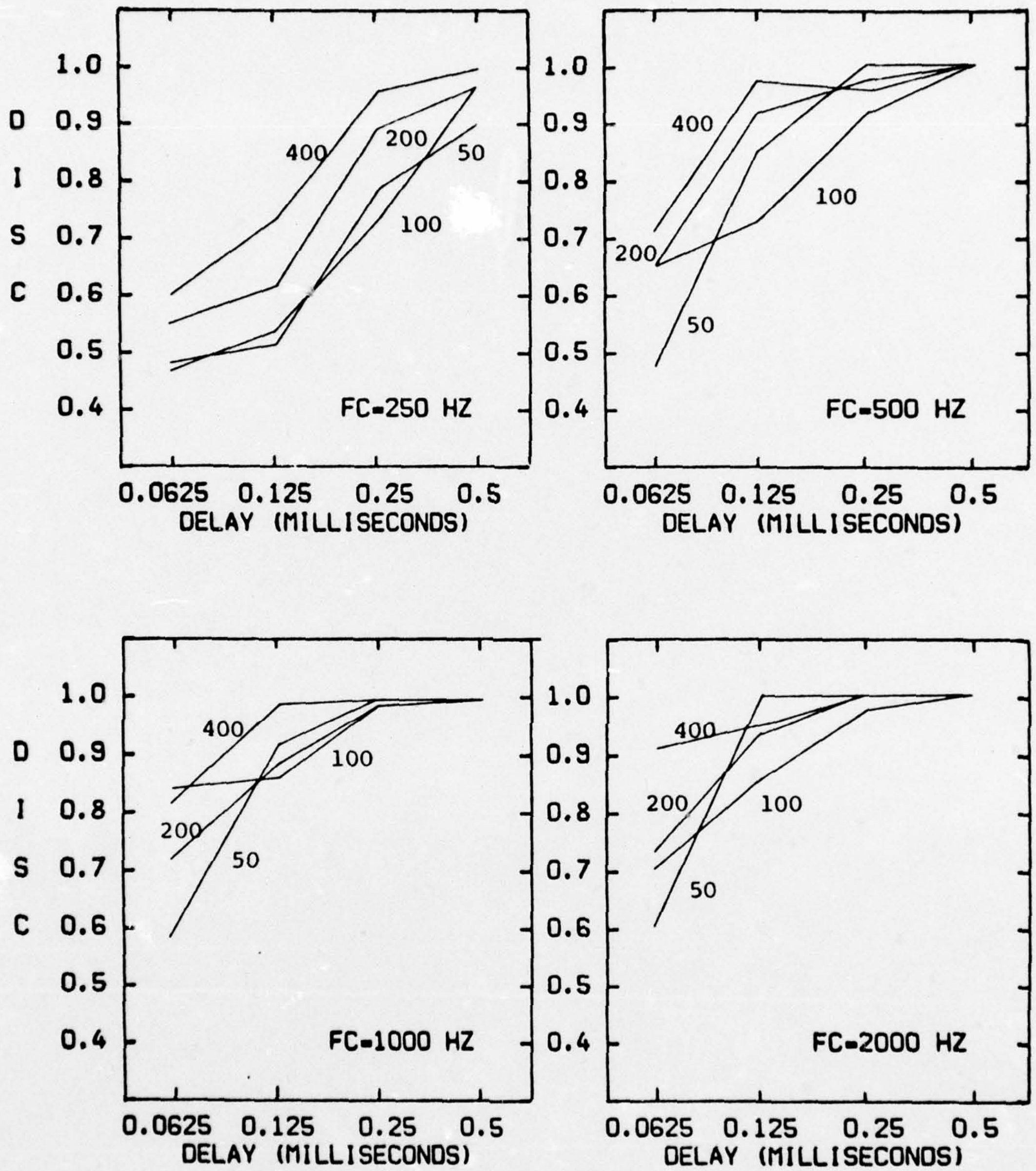


Figure 13. Discriminability of Subject WG between an impulse and phase modified impulses plotted as a function of delay with bandwidth as parameter.

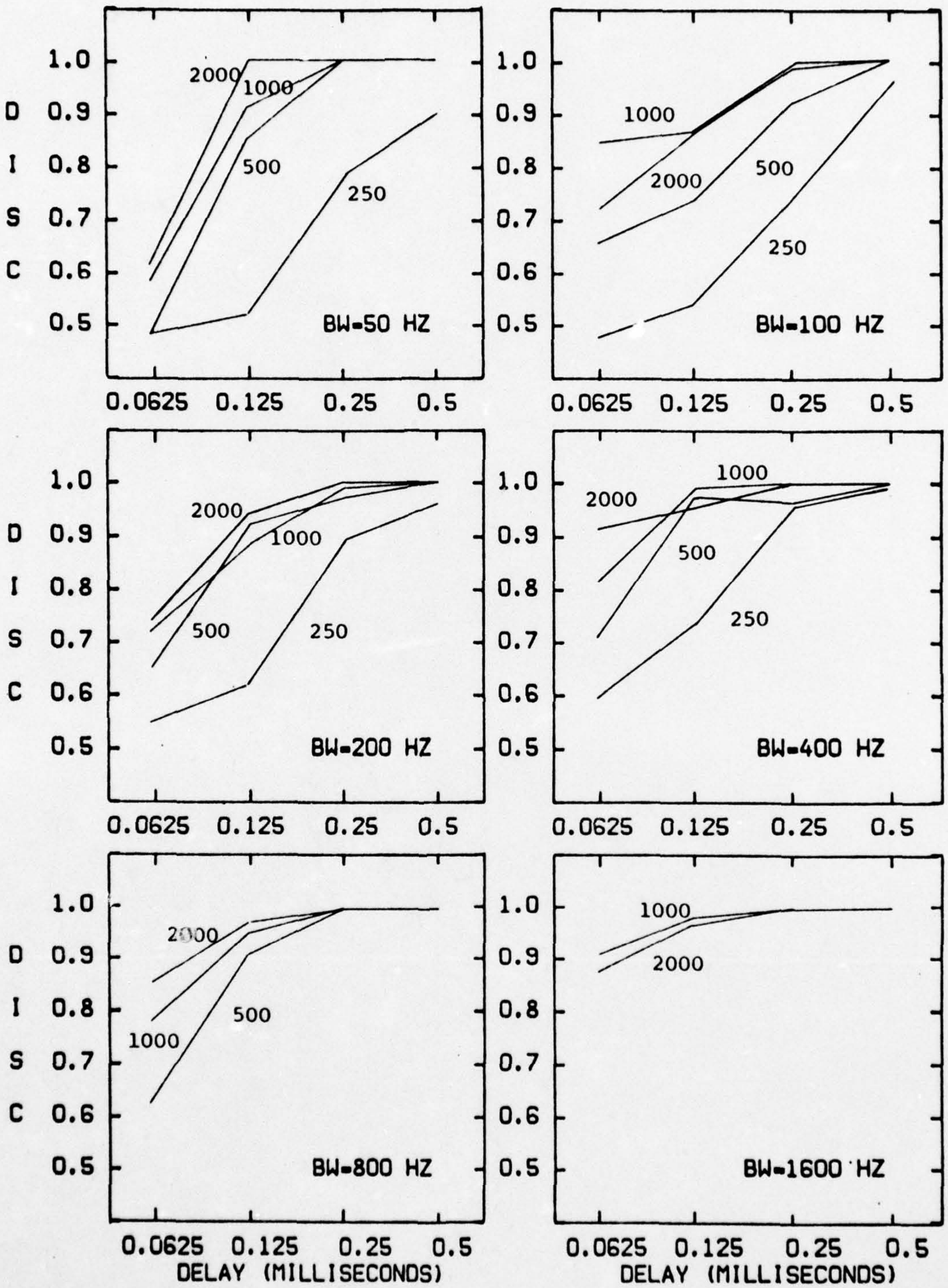


Figure 14. Discriminability of Subject WG between an impulse and phase modified impulses plotted as a function of delay with center frequency as parameter.

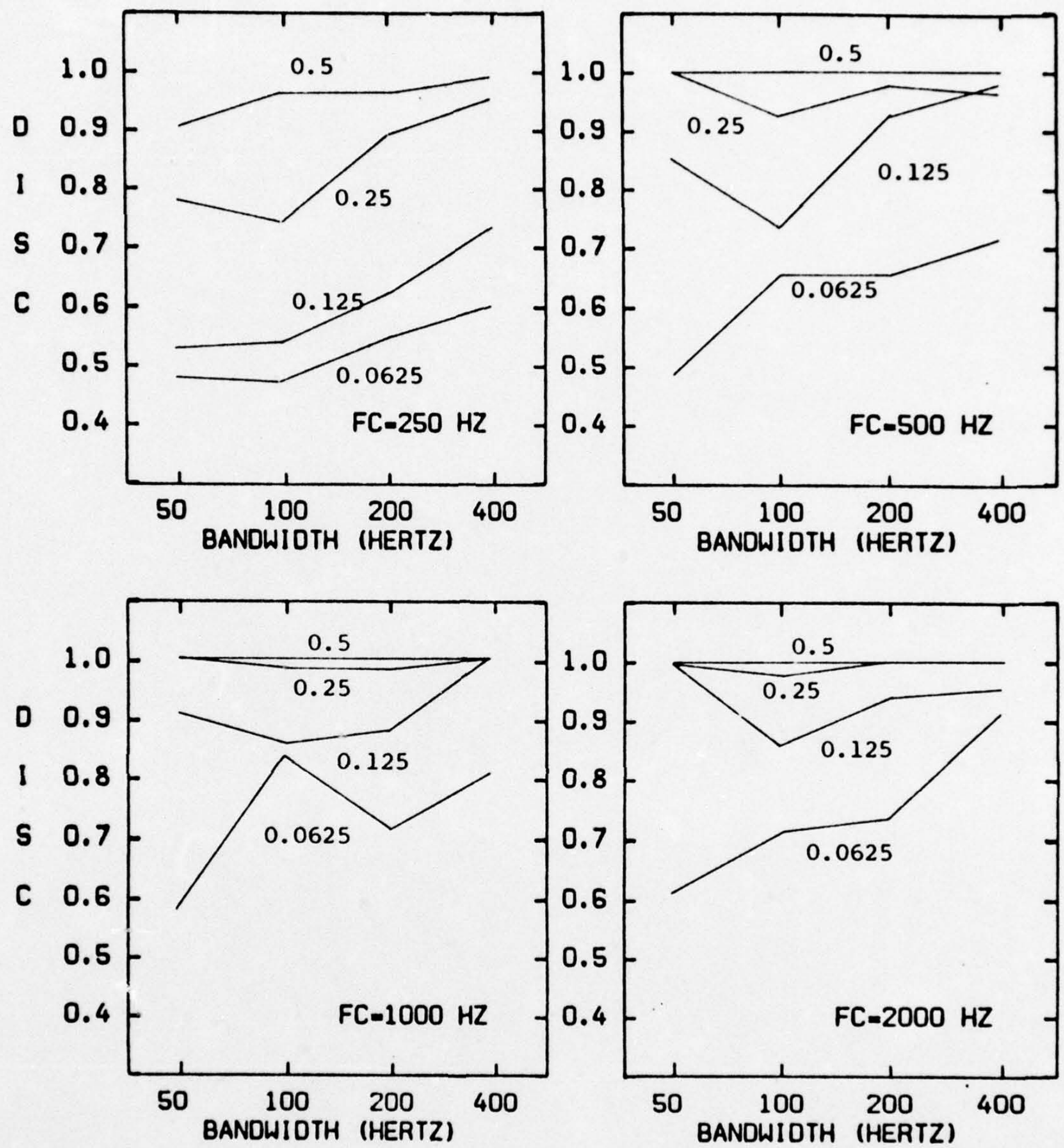


Figure 15. Discriminability of Subject WG between an impulse and phase modified impulses plotted as a function of bandwidth with delay as parameter.

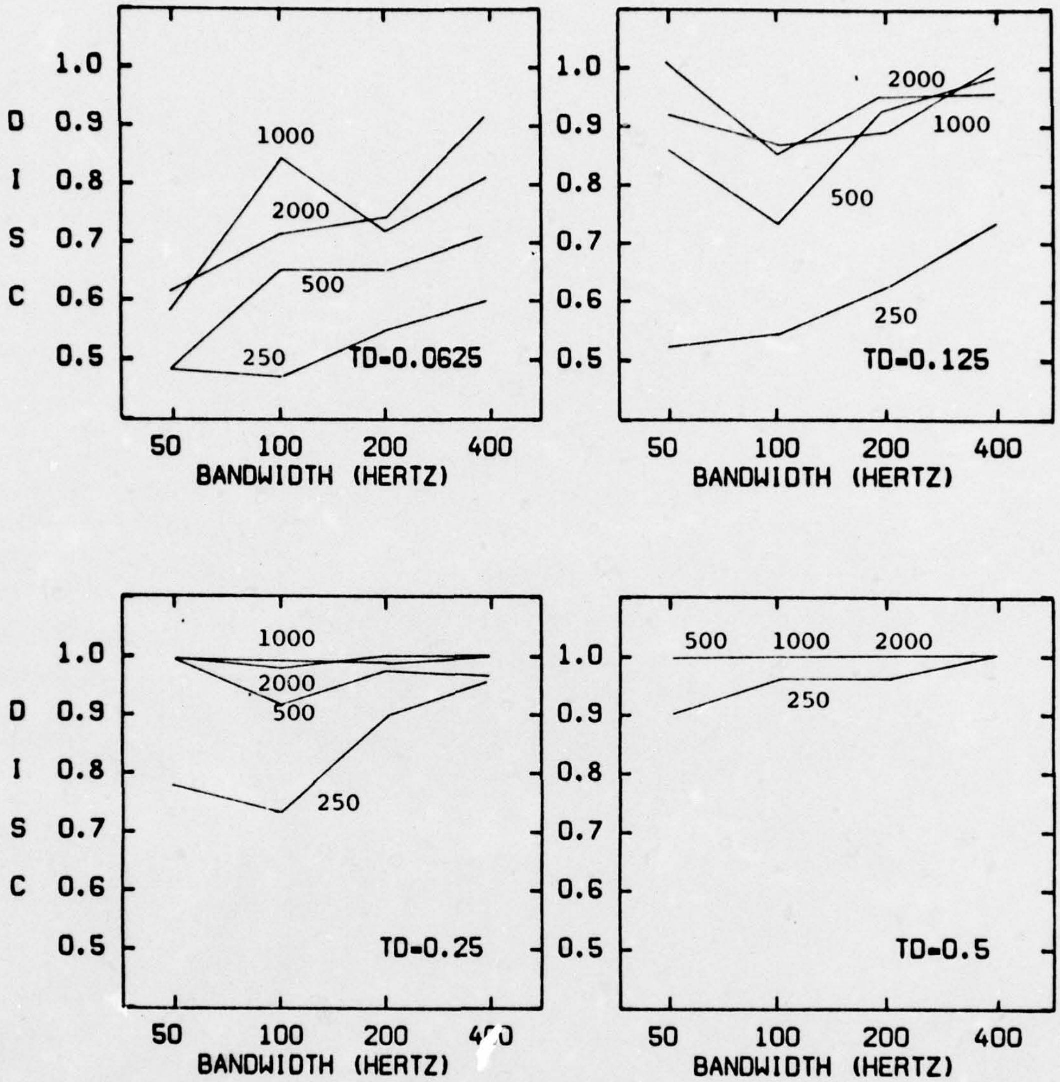


Figure 16. Discriminability of Subject WG between an impulse and phase modified impulses plotted as a function of bandwidth with center frequency as parameter.

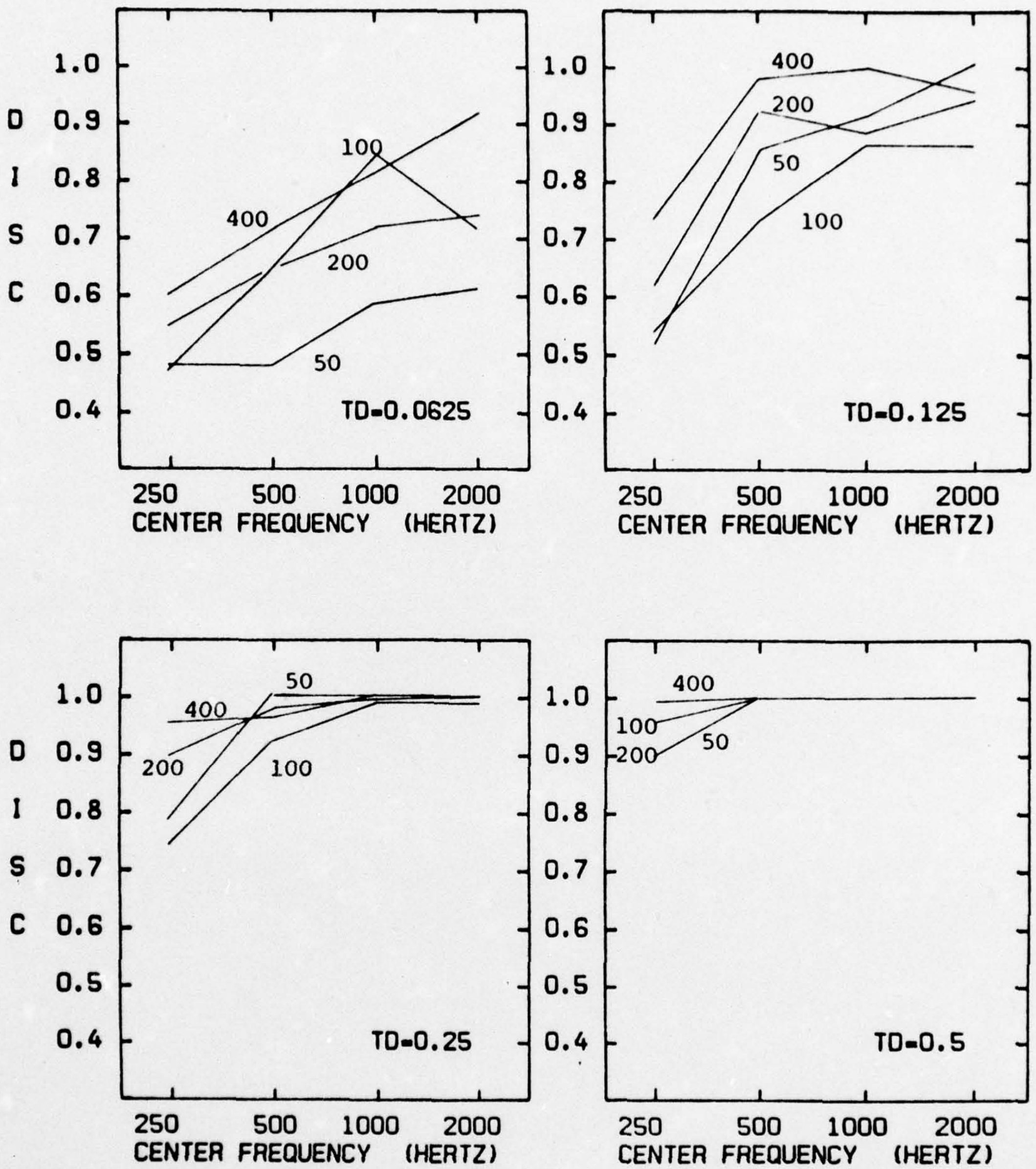


Figure 17. Discriminability of Subject WG between an impulse and phase modified impulses plotted as a function of center frequency with bandwidth as parameter.

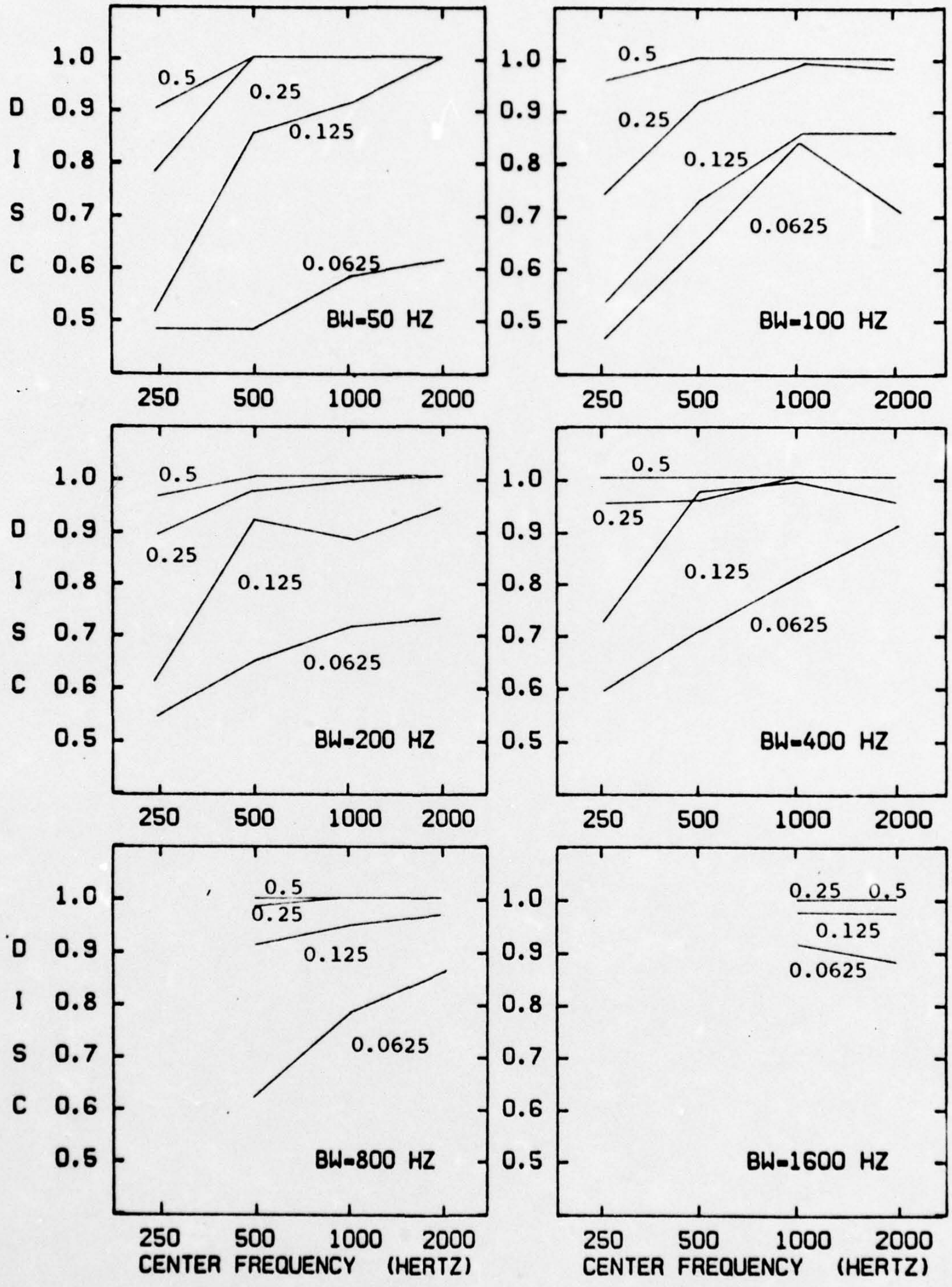


Figure 18. Discriminability of Subject WG between an impulse and phase modified impulses plotted as a function of center frequency with delay as parameter.

Discussion

Results for the 500 Hertz center frequency condition shown in Figure 13 should be noted to agree with those of the three subjects first tested for the same phase condition shown in Figure 8. It appears that the author's phase sensitivity is slightly more acute in that he is able to achieve 0.75 discrimination between 0.0625 and 0.125 milliseconds of delay for the 500 Hertz center frequency condition as opposed to 0.125 and 0.25 milliseconds as required by the other three subjects. This is no doubt due to his greater familiarity with the stimuli and all of their nuances of quality.

The data indicates that discriminability is mildly, yet significantly dependent on bandwidth. Discriminability is perfect for certain values of center frequency, bandwidth and delay. Discriminability would fall to chance, however, if bandwidth were reduced to zero. One would expect that discriminability monotonically increases with bandwidth. The values selected for bandwidth as the independent variable do not allow the observance of the monotonic increase through a "knee" or "point of inflection" in the function. The discriminability function appears to have reached an upper plateau which is asymptotically decreasing for the selected values of bandwidth. The displaced position of the 50 Hertz bandwidth curve in Figure 13 and 17 from an otherwise monotonic ordering is probably due to its relative closeness to a point of inflection in the function. The experiment has thus demonstrated that the range of interest for bandwidth values is less than 100

Hertz. The ubiquitous critical band phenomenon in psychophysical research may account for the leveling off of discriminability for bandwidths greater than 100 Hertz.

Envisioning the topology of a function of three independent variables is difficult at best. Inasmuch as the effect of bandwidth for the values selected has significantly less effect on discriminability than center frequency or delay, the relative effects of center frequency and delay may be observed by averaging the data across bandwidth and plotting discrimination as a parametric function of center frequency and delay as illustrated in Figure 19. The figure was generated by first drawing smooth curves by eye through the data plotted as functions of center frequency with delay as parameter and as functions of delay with center frequency as parameter. Iso-contour curves of the parametric function were drawn by eye through points generated from orthogonal projections of points from the previously drawn curves.

The contours illustrated in Figure 19 appear to be very well behaved. The contours would appear to asymptotically converge with three further halfings of center frequency at approximately 31 Hertz or the lower frequency limit of hearing. One may observe from Figure 19 that for bandwidths greater than 100 Hertz and for a delay of 0.0625 milliseconds that a discrimination score of 0.75 is achieved at approximately 1000 Hertz center frequency. For a center frequency of 250 Hertz, the 0.75 score is achieved at approximately 0.25 milliseconds.

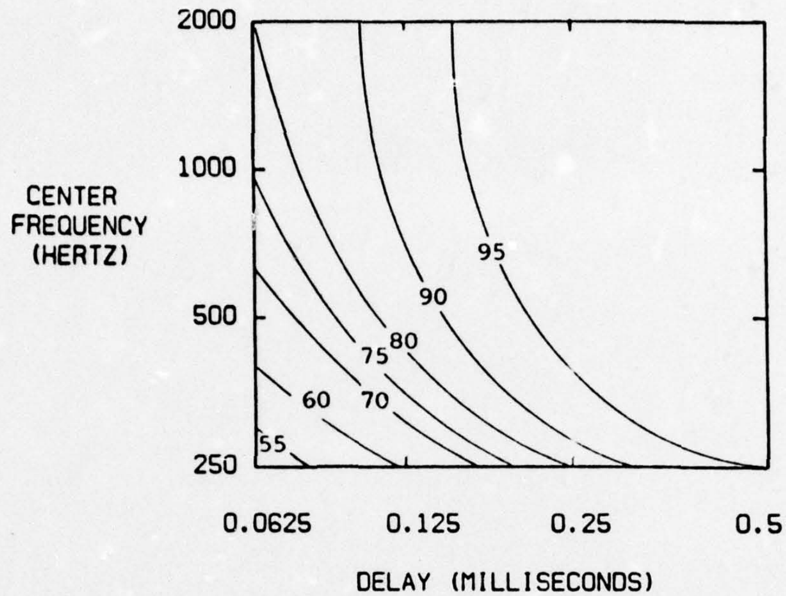


Figure 19. Relative effect of center frequency and delay on discriminability.

In summary, a series of experiments was conducted to determine the discriminability between a standard impulse and impulses in which a band of frequency components are delayed. The parameters of interest were intensity, center frequency, bandwidth and delay. Discriminability was shown to be highly dependent on intensity. At a sensation level of 80 dB, discriminability was shown to be mildly, yet significantly dependent on bandwidth for bandwidths greater than 100 Hertz. The contribution of center frequency and delay to discriminability was shown to be inversely related. Discrimination scores of 0.75 were achieved for center frequencies between 250 and 500 Hertz at delays between 0.0625 and 0.125 milliseconds.

The phase discriminability function was at first hoped to be determined with high resolution by conducting many tests in which the parameters of interest could be adjusted over a wide range in relatively small increments. Initial attempts by the author in determining the range of interest of the experimental variables was frustrated by the time required to measure accurately a point in the function, and the apparently large variability of the function. Only an approximate estimate of the phase discriminability function must consequently suffice.

CHAPTER VII

SENSITIVITY TO DISPERSION IN SPEECH

Direct assessment of the effect of dispersion on continuous speech is difficult because of the short-time spectral complexity of speech and the constant shift in short-time spectra during production of the speech signal. The difficulty in theoretically estimating the discriminability of dispersion based on short-time spectral features has been not so much due to the inherent complexity of the spectra itself as to lack of information concerning the sensitivity of the ear to dispersion in general. One purpose of the experiments discussed in Chapter VI was to learn more about phase sensitivity so as to make possible generalizations about phase modified continuous speech.

Speech consists of strings of concatenated phonemes whose short-time spectra are modified by such factors as articulation, coarticulation, the emotional state of the speaker, etc. The constant shift in the short-time spectra of speech makes assignment of a phase discriminability statistic to continuous speech difficult because of the varying degree of dependence of phase discriminability on each successive phoneme. For example, one would expect less sensitivity to dispersion in fricatives because of their already noise-like characteristics than in plosives with their sudden and relatively simultaneous onset of frequency components.

One might expect that dispersion in a spectral region in which speech has statistically little energy would result in smaller overall subjective effects than would dispersion in a spectral region with a statistically higher energy concentration. A tacit assumption of this expectation is that the perception of quality change is due to cumulative effects for each phoneme over the phonetic utterance. One could hypothesize, consequently, that a general discriminability statistic for continuous speech might be derived from a weighted sum of the individual effects of dispersion on the constituent phonemes. On the other hand, quality changes may become readily apparent if only one or two relatively "sensitive" phonemes are are dispersed only slightly.

Little is known of the ear's phase discriminability ability outside of the well known fact that excessive dispersion in speech results in chirp-like sounds. The intent of the experiments discussed in this chapter is to further investigate the phase discriminability phenomenon as it pertains to speech stimuli.

Sensitivity to Dispersion in Phonemes

The major intent of the experiment to be discussed in this section is to obtain a set of psychophysical data which is indicative of human discriminability of dispersion of the type introduced in Chapter IV for various classes of phonemes. Such data should be of interest in directly probing the speech discriminability mechanism of hearing inasmuch as various test results derived from psychophysical measurements using different classes of phonemic stimuli can be analyzed in relation to the

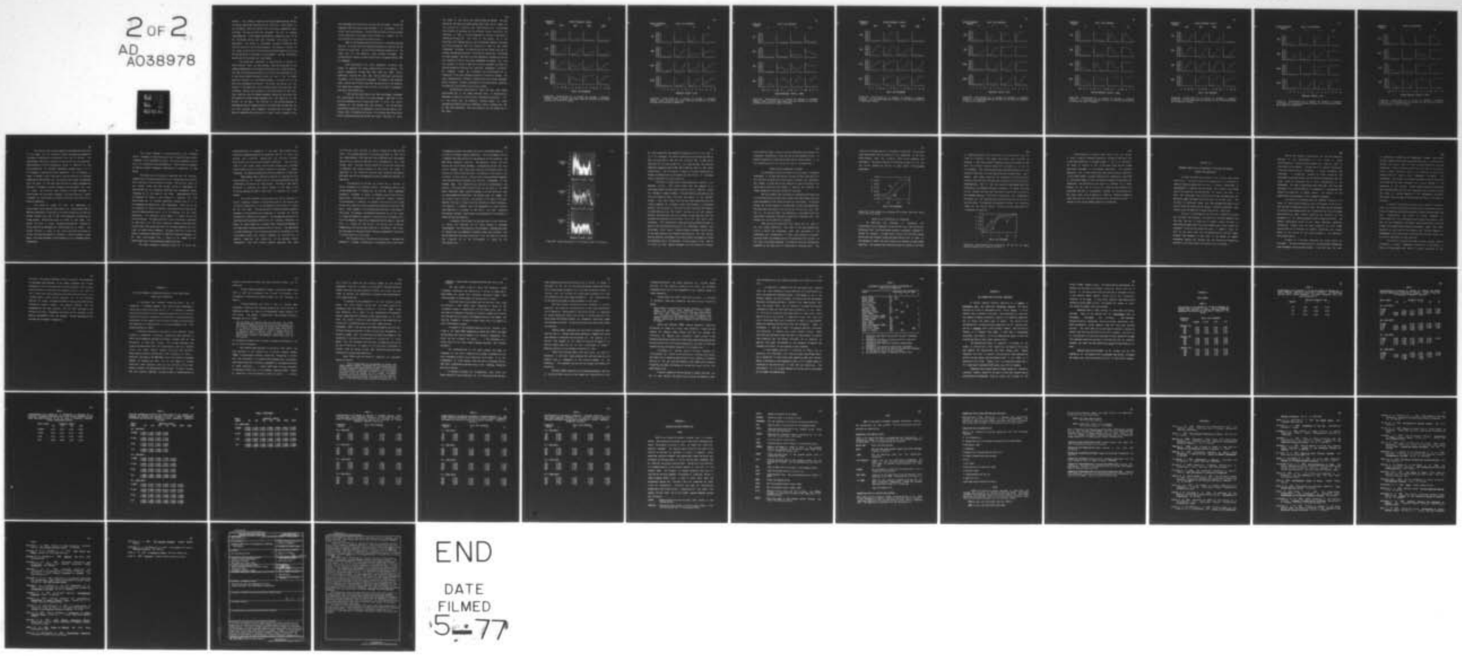
stimuli. The data should also be applicable to indirectly estimating human discriminability of dispersion in continuous speech. The phonemes of choice used as stimuli in the current experiment were selected to represent three classes of phonemes and consequently included the neutral vowel /ə/, the plosive /t/ and the fricative /f/. These sounds were obtained from digital recordings as discussed in Chapter IV. The vowel stimulus was obtained from a recording of the vowel spoken in isolation whereas the piosive and fricative were extracted from recordings of single words spoken in isolation. The phoneme extraction process consisted of displaying the signal from which a phoneme was to be extracted on a storage oscilloscope so as to determine the beginning and ending sample of the phoneme segment. Auditory playback was used to insure the quality of the isolated segment. A trapezoidal window with a rise and decay time of 20 milliseconds was used to isolate the segment after which the segment was copied to disk as a new audio file for later reference.

The phonemes were presented in the discrimination tests during the course of the experiment at their normal conversational speech level. The levels were determined in the following way: Stimulus thresholds for the two words spoken in isolation were determined by the method of adjustment. The author conducted a series of tests with himself as subject in which he alternately decreased the setting of the signal attenuator in the audio playback circuit until the stimuli just became inaudible and then increased the setting of the attenuator setting until the stimuli just became

AD-A038 978 UTAH UNIV SALT LAKE CITY DEPT OF COMPUTER SCIENCE F/G 6/16
MONAURAL SENSITIVITY TO DISPERSION IN IMPULSES AND SPEECH. (U)
DEC 75 W H GREER DAHC15-73-C-0363

UNCLASSIFIED UTEC-CSC-75-142 NL

2 OF 2
AD A038978



END
DATE
FILMED
5-77

audible. The average attenuator setting was approximately 90 dB. The author determined that both stimuli were fully intelligible at an attenuator setting of 50 dB, which is to say, a sensation level of 40 dB. The level at which the attenuator was set to produce "conversational level" speech was 20 dB for a sensation level of 70 dB. The 20 dB setting was used throughout the course of the experiment. The method of adjustment was used to determine the relative levels of the isolated phonemes. The sensation levels of the vowel, plosive and fricative were determined respectively to be 70, 32 and 35 dB in reference to the attenuator setting previously determined for conversational level speech.

The experiment consisted of administering a series of discrimination tests for each phoneme. Each test consisted of the random presentation of 16 unique stimulus pairs for discrimination. The set of stimulus pairs peculiar to a given test consisted of one of the three standard phonemes followed by one of the 16 phase modified variants of the standard phoneme. The stimuli of all 16 pairs were consequently different. The author, serving as his own subject in the experiment, was of necessity aware that each pair was different. Control stimulus pairs in which the stimuli of the pair were identical were not deemed necessary inasmuch as the author was motivated not to make random responses as to the equality of the stimuli of the pair. The validity of the testing paradigm is attested by the well behaved nature of the data when plotted and by low test variance when repeated. The author set for himself the task of responding that the stimuli of a pair were different only

when persuaded that the stimuli actually were different. The author responded that the stimuli were the same if no difference in the stimuli could be detected. A discrimination score of 0.5 was chosen to be indicative of the author's discriminability inasmuch as all the stimulus pairs contained different stimuli.

Each of the 16 phase modified variants of a standard phoneme peculiar to a given test was randomly presented 10 times during the course of a given test. The pairs were presented monaurally to the right ear at the sensation levels already discussed. The intra-stimulus interval of each stimulus pair was approximately 500 milliseconds.

Four values each of the three independent variables were chosen to generate 64 unique phase conditions for testing. The center frequencies included 250, 500, 1000 and 2000 Hertz; bandwidths included 50, 100, 200, 400 and 800 Hertz; and delays included 2, 4, 8, 16 and 32 milliseconds. The phase of the standard vowel, plosive and fricative were modified according to the schedule just described to generate a set of stimuli to be used in subsequent testing of discriminability.

A test series consisting of four tests as already discussed was administered for each of the three phonemes. Each test in the series represented a set of phase conditions in which the center frequency of the delayed band was constant. The 16 pairings represented by each test corresponded to the unique pairings of four values each of bandwidth and delay. All 64 phase conditions could thus be tested by administering the four tests. The order in which

the tests of the series was administered was random. The test series for the vowel was administered three times while those for the plosive and fricative were administered twice. The average discriminability estimates for the different phase conditions are tabulated in Table 7, 8 and 9 (Appendix C) and are illustrated in Figures 20 through 28. The intent of the author, as in the experiment with impulse stimuli, was to determine trends in the data by first plotting the data as functions of each of the three independent variables, to characterize the the trends, and then to discuss the characterizations. Three sets of curves are illustrated for each phoneme. Each set of curves represents the data as plotted as a function of each of the three independent variables. The rows and columns of curves in each set of curves represent the data for various parametric values of the remaining independent variables. For example, Figure 20 illustrates the discriminability of dispersion in the vowel phoneme plotted as a function of delay. The rows parametrically represent bandwidth and the columns represent center frequency. Effects of parametric variations may be observed by comparing curves across rows or down columns.

The majority of the curves in Figure 20 show that phase discriminability in the phase modified vowel is monotonically dependent on delay for a given center frequency and bandwidth. All of the curves show the monotonic increase except for those representing phase conditions of 500 Hertz center frequency and 400 or 800 Hertz bandwidth. The significance of the two exceptions is not clear.

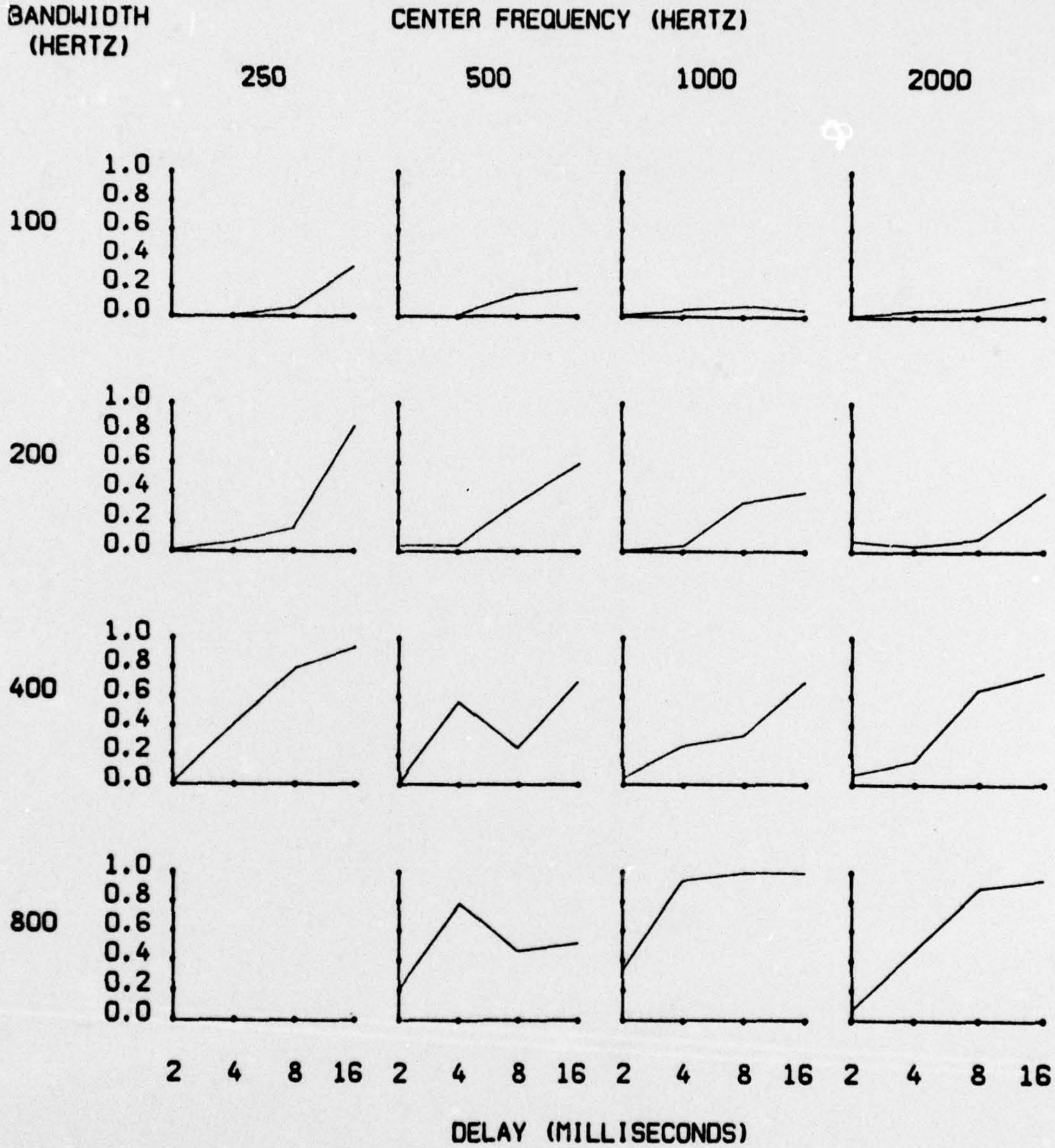


Figure 20. Discriminability of Subject WG between a standard neutral vowel stimulus and phase modified variants of the vowel plotted as a function of delay.

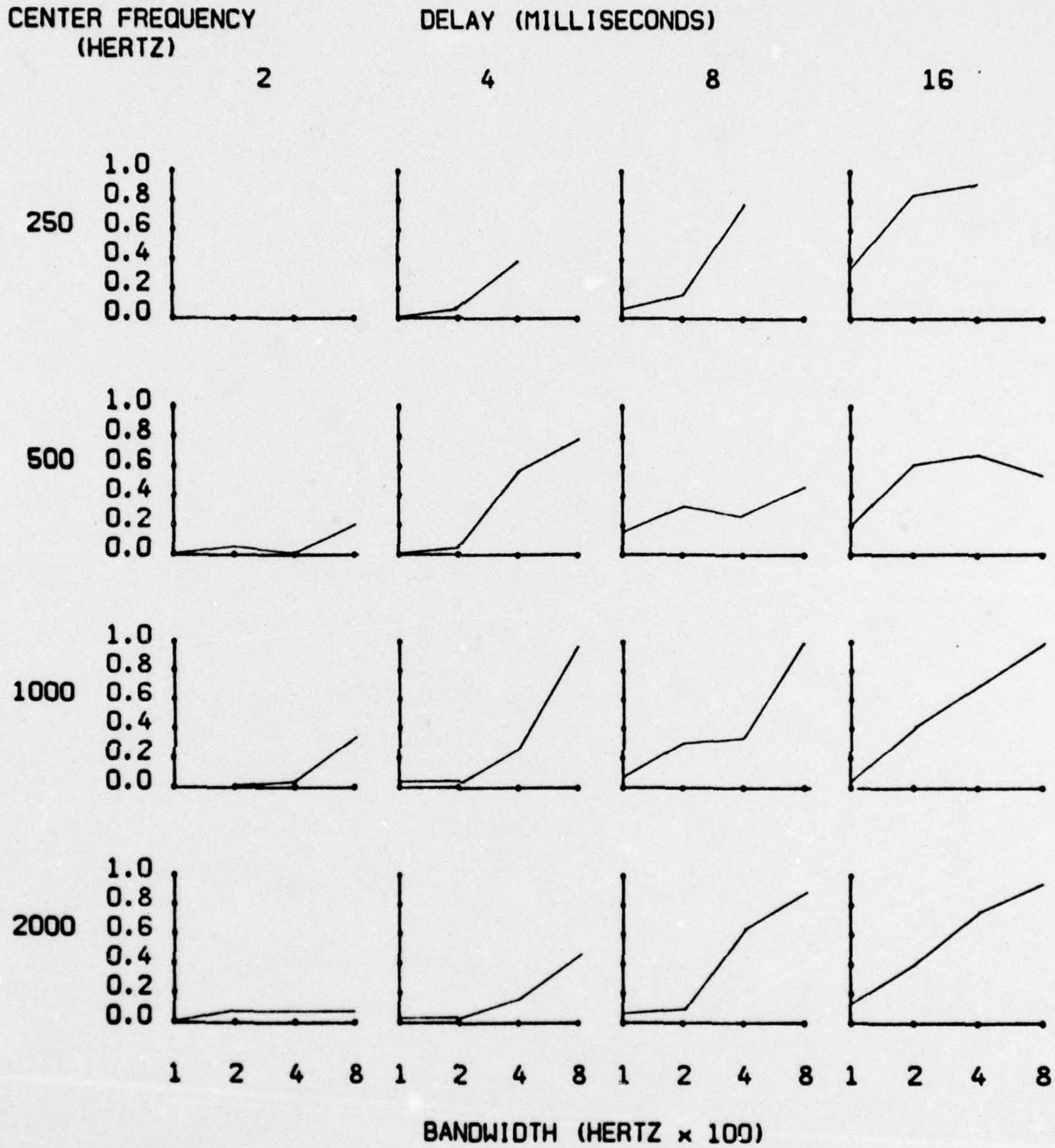


Figure 21. Discriminability of Subject WG between a standard neutral vowel stimulus and phase modified variants of the vowel plotted as a function of bandwidth.

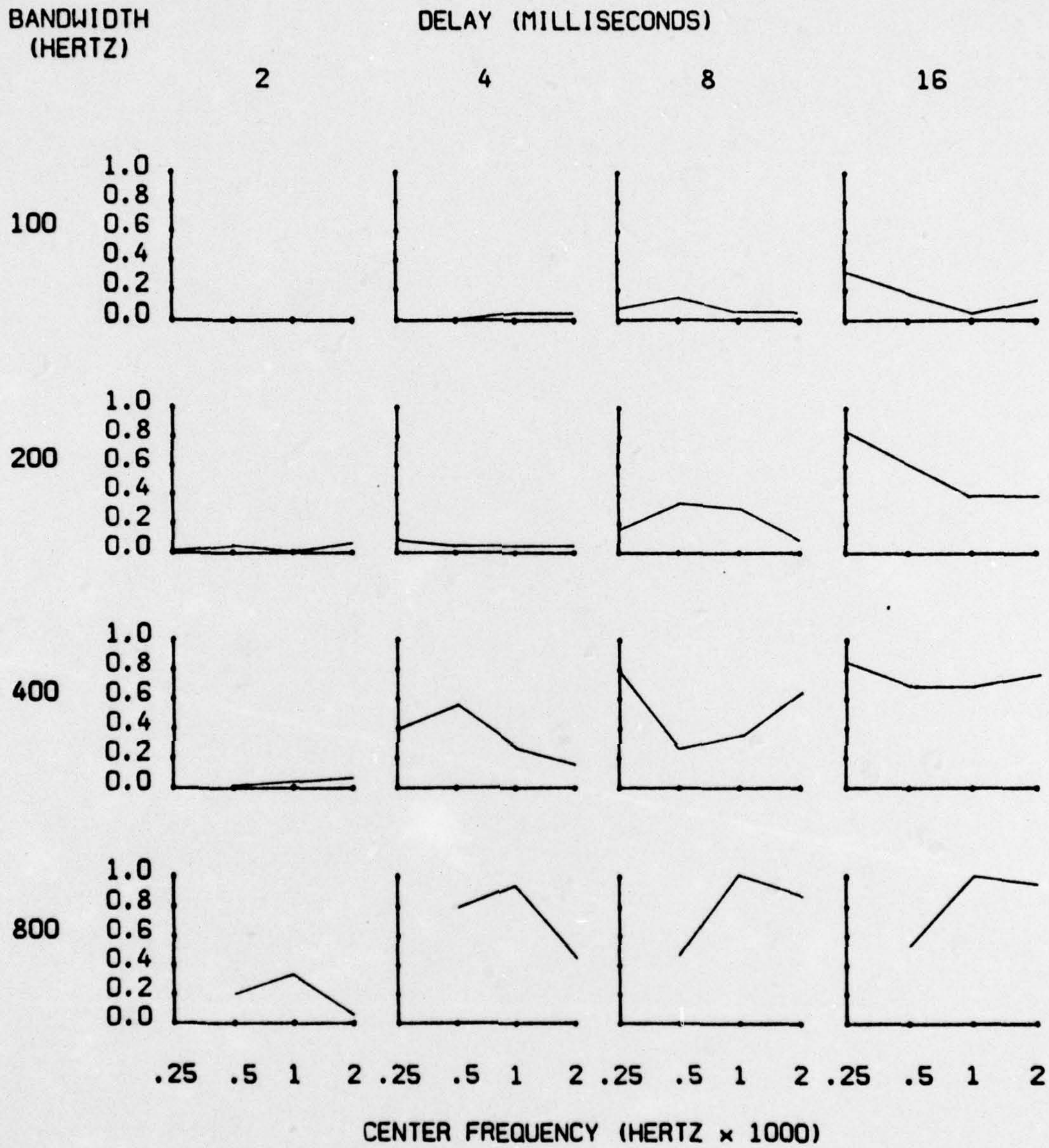


Figure 22. Discriminability of Subject WG between a standard neutral vowel stimulus and phase modified variants of the vowel plotted as a function of center frequency.

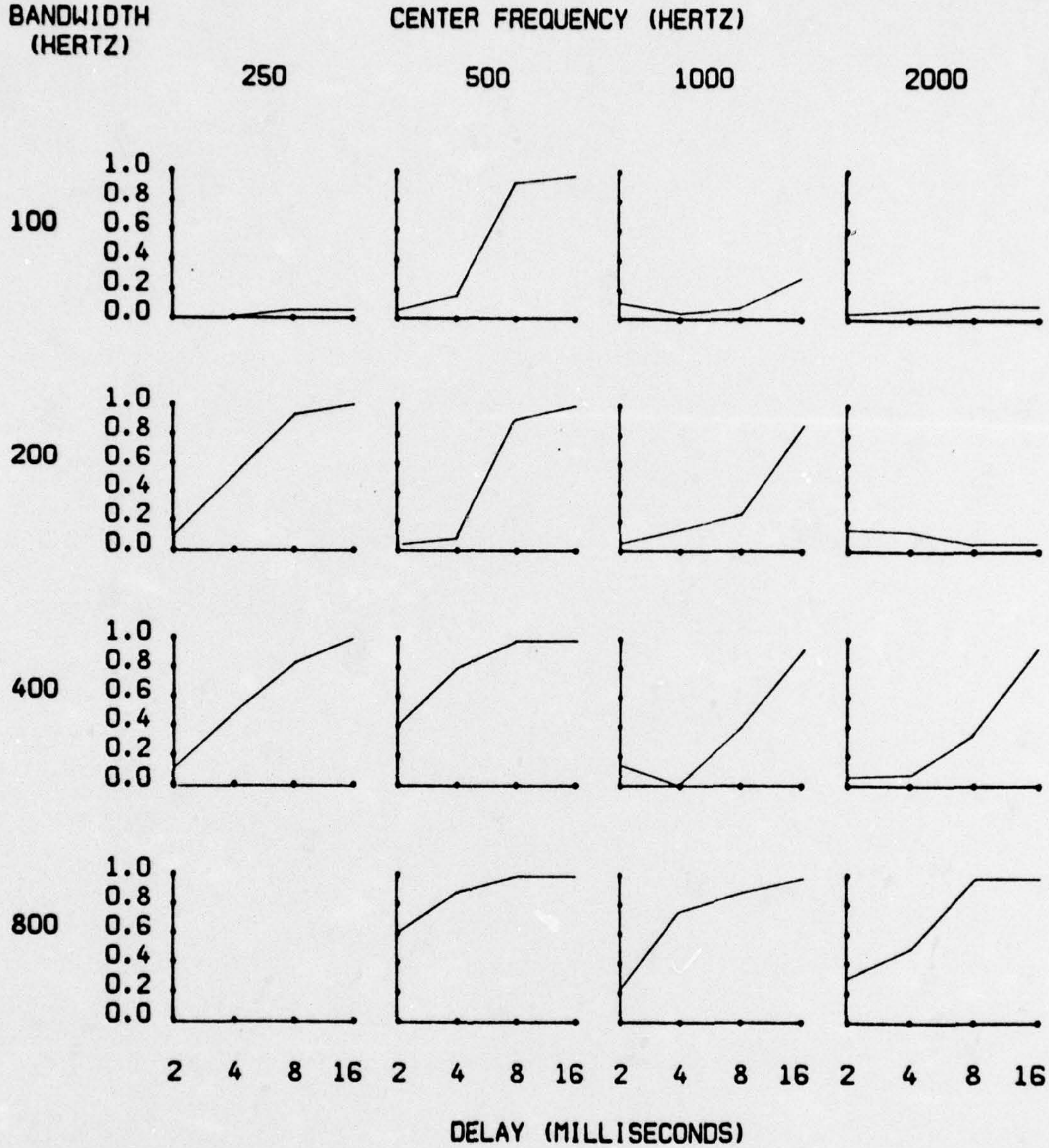


Figure 23. Discriminability of Subject WG between a standard plosive, /t/, and phase modified variants of the plosive plotted as a function of delay.

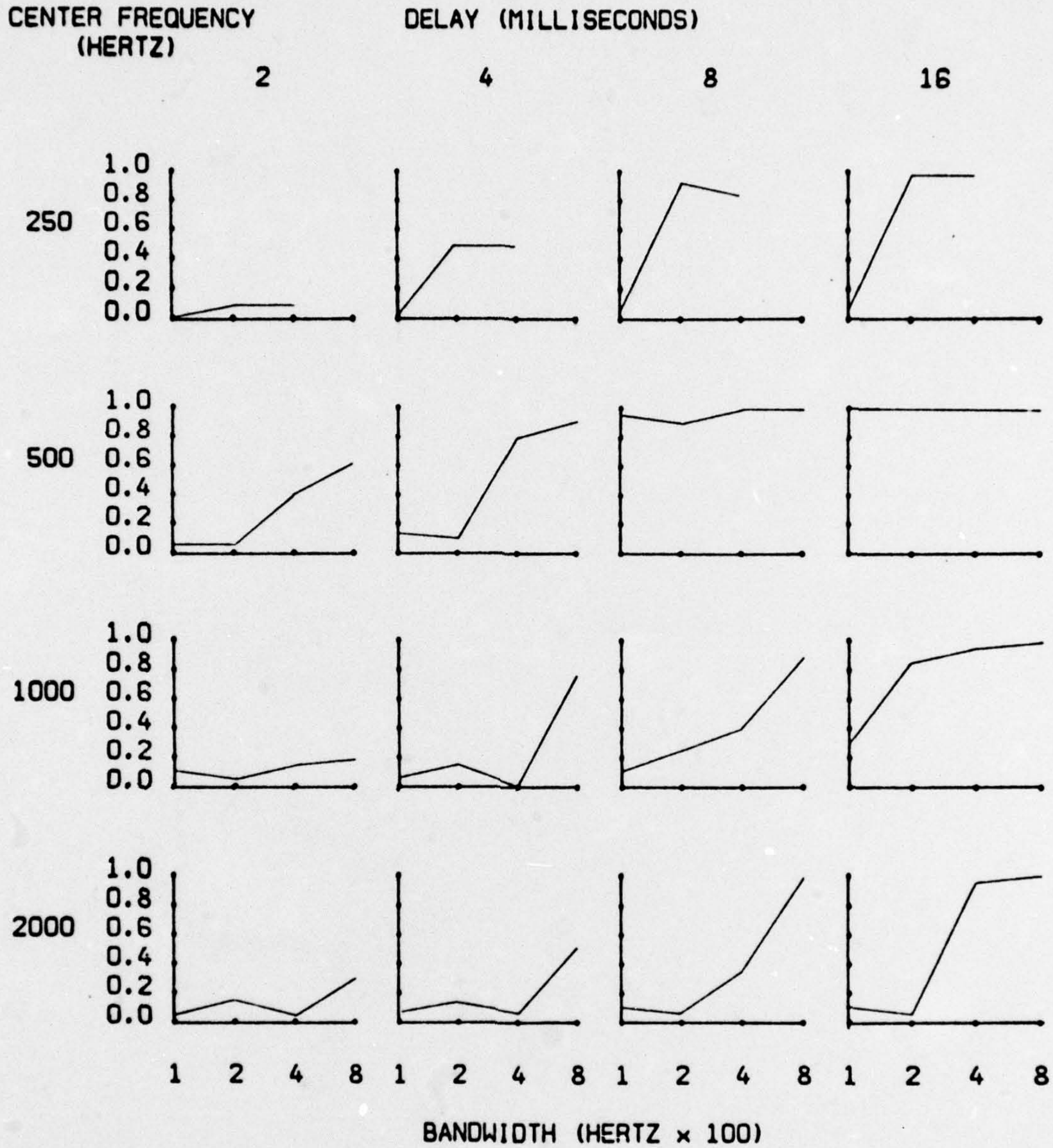


Figure 24. Discriminability of Subject WG between a standard plosive, /t/, and phase modified variants of the plosive plotted as a function of bandwidth.

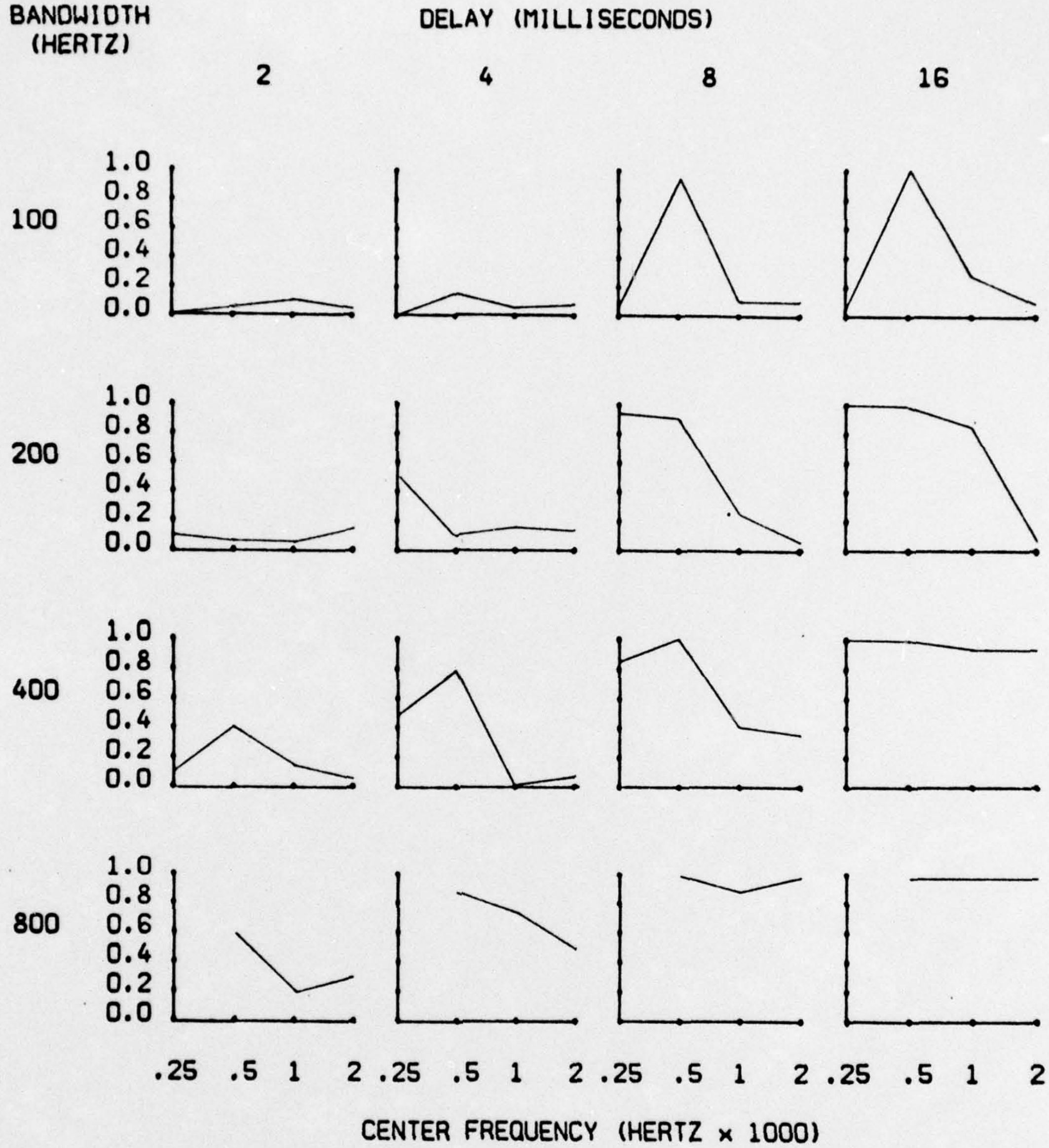


Figure 25. Discriminability of Subject WG between a standard plosive, /t/, and phase modified variants of the plosive plotted as a function of center frequency.

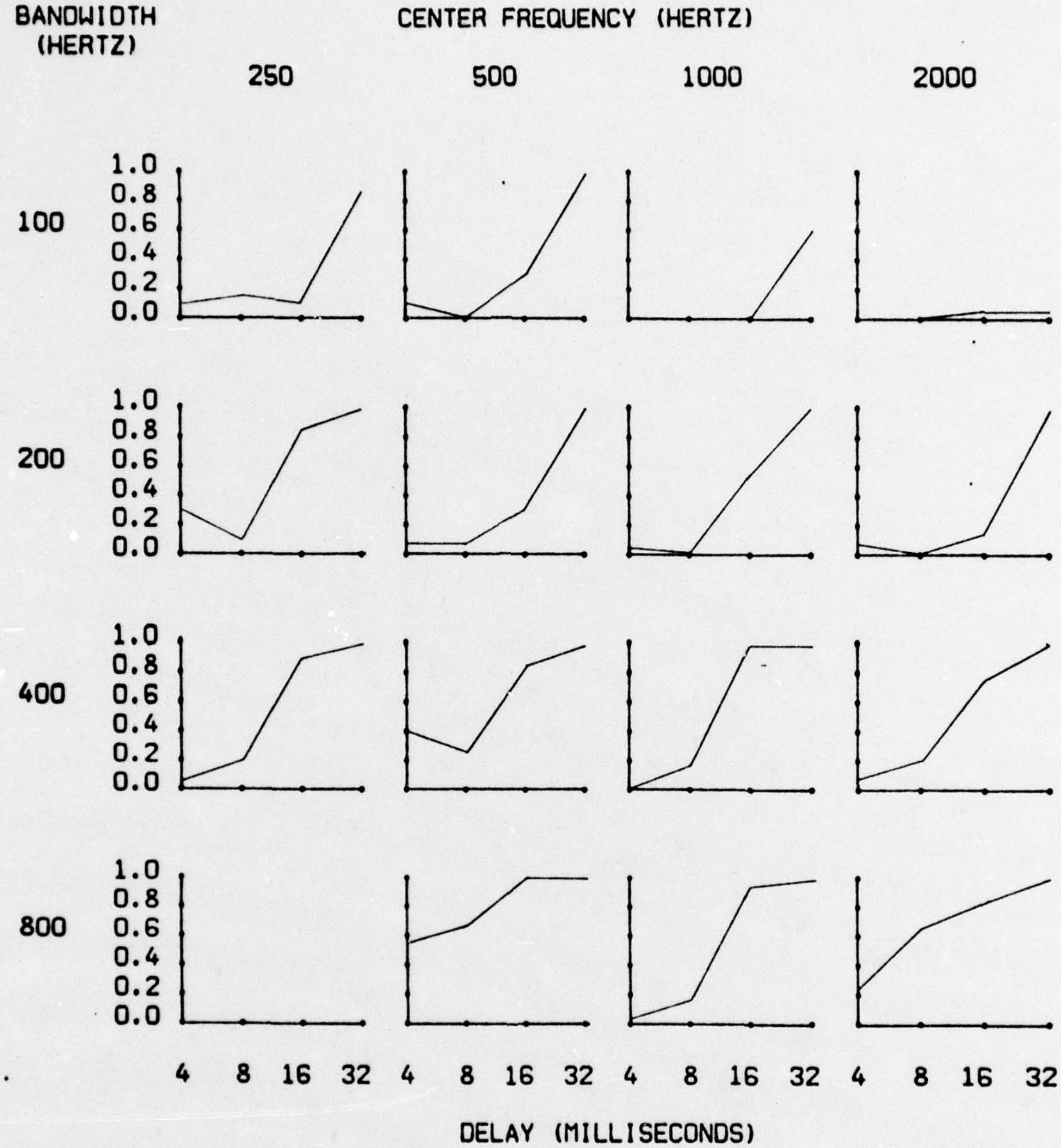


Figure 26. Discriminability of Subject WG between a standard fricative, /f/, and phase modified variants of the fricative plotted as a function of delay.

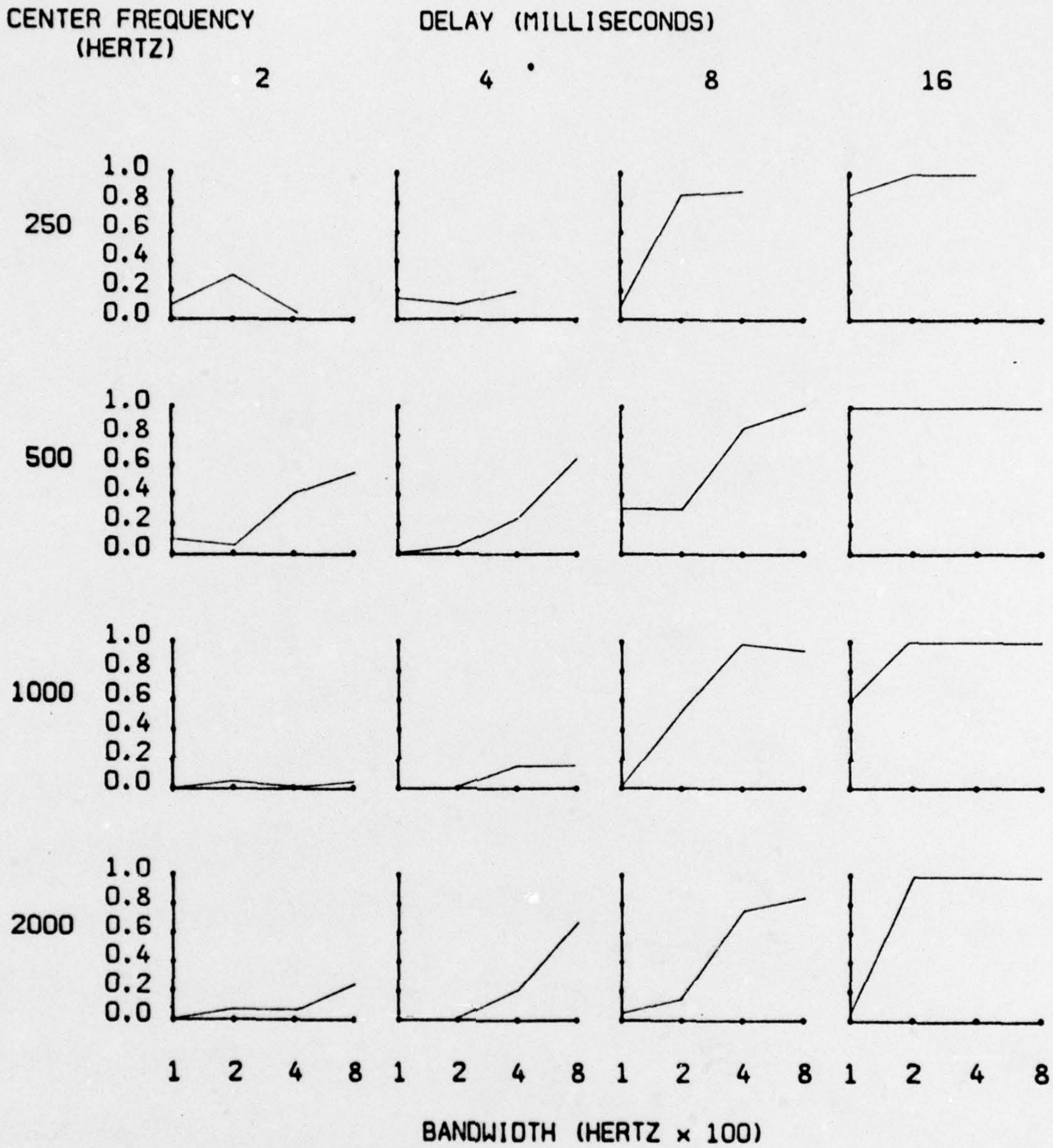


Figure 27. Discriminability of Subject WG between a standard fricative, /f/, and phase modified variants of the fricative plotted as a function of bandwidth.

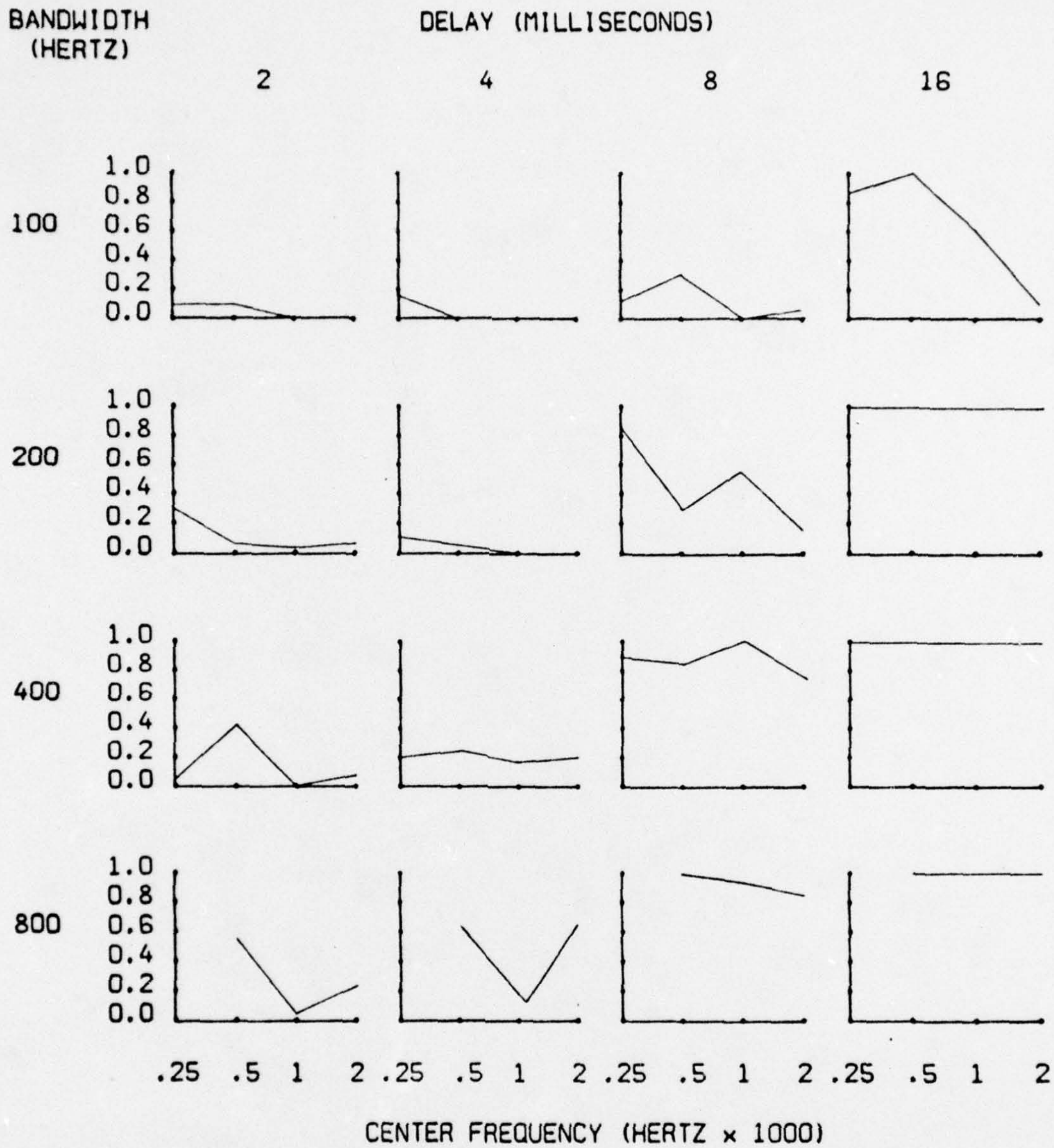


Figure 28. Discriminability of Subject WG between a standard fricative, /f/, and phase modified variants of the fricative plotted as a function of center frequency.

The locus of the curves is generally displaced more and more to the upper left as columns of curves, representing parametric increases in bandwidth, are considered from top to bottom. The displacement implies an increase in discriminability with bandwidth. Approximately 8 to 16 milliseconds of delay is required for 0.5 discriminability at narrower bandwidths whereas only 2 to 4 milliseconds is required at wider bandwidths. The difference in range of delays is most likely due to the contribution of increased bandwidth on discriminability. The locus of the curves is displaced more and more to the lower right as rows of curves, representing parametric increases in center frequency, are considered from left to right. The shifting locus implies a slight decrease in discriminability with center frequency. The contribution of center frequency to discriminability appears to be much less than that of delay or bandwidth.

The curves of Figure 21 show the dependence of discriminability on bandwidth more explicitly. Approximately 200 to 800 Hertz bandwidth is required for 0.5 discriminability for shorter delays whereas only 100 to 400 Hertz bandwidth is required for longer delays. The shifting locus of curves to the upper left across the rows of curves, representing parametric increases in delay, implies the dependence of discriminability on delay. The slight shifting of locus to the lower right down the column of curves, representing parametric increases in center frequency, shows again the slight decrease in discriminability with increased center frequencies.

The slight decrease in discriminability with increased center frequency is shown more explicitly in Figure 22 where center frequency is the independent variable. The relation appears to hold only for bandwidths less than 800 Hertz. Discriminability appears to peak at a center frequency of 1000 Hertz for bandwidths of 800 Hertz.

The majority of the curves of Figures 23 and 26 indicate, respectively, that phase discriminability in phase modified plosives and fricatives is also a function of delay. The shifting locus of the curves along rows and columns implies a dependence of discriminability on increasing bandwidth and decreasing center frequency as in the case of the vowel. Approximately 4 to 8 milliseconds of delay in the plosive is required for 0.5 discriminability at narrower bandwidths whereas only 2 to 4 milliseconds are required for wider bandwidths. From 16 to 32 milliseconds of delay is required in the case of the fricative for narrower bandwidths and from 4 to 8 milliseconds for the wider bandwidths. Two anomalous curves appear in Figure 23 for the plosive phoneme. The phase condition in which the delay is centered at 500 Hertz with 100 Hertz bandwidth appears to be much more discriminable than phase conditions with the same bandwidth but with lower or higher center frequency. The phase condition in which the delay is centered at 2000 Hertz with 200 Hertz bandwidth appears to be less discriminable than it should be as judged by the discriminability of the surrounding phase conditions.

The same discussion regarding Figure 21 in which the

discriminability of dispersion in the vowel was plotted as a function of bandwidth applies to Figures 24 and 27 in which the plosive and fricative, respectively, are similarly plotted. Discriminability is definitely related to bandwidth. The shifting locus of the curves along rows and down columns imply increased discriminability with an increase in delay or a decrease in center frequency. No anomalous phase conditions are noted as in Figure 23.

Approximately 200 to 400 Hertz bandwidth is required for 0.5 discriminability for shorter delay in the plosive whereas only 100 to 200 Hertz is required for longer delays. From 400 to 800 Hertz bandwidth is required for shorter delays in the case of the fricative whereas only 100 to 400 Hertz is required for longer delays.

The slight decrease in discriminability with an increase in center frequency as shown in Figures 25 and 28 for the plosive and fricative, respectively, is not as apparent as that shown in Figure 22 for the vowel. The plosive data in Figure 25 shows a general decrease in discriminability of dispersion in plosives for center frequencies of 500 Hertz and greater. The decrease may be partially mediated by the decreased contribution of a constant bandwidth of delay to discriminability at higher center frequencies because of the logarithmic frequency characteristic of the ear. The hypothesis is weakly supported in the following way: Each of the spectra of the three speech sounds were visually searched for similarities of spectral magnitude and composition at two adjacent center frequencies. Only the fricative phoneme magnitude had such

similarities which occurred at center frequencies of 250 and 500 Hertz. The average magnitude of the energy centered at 500 Hertz was approximately 2 dB less than that at 250 Hertz but the greater sensitivity of hearing at 500 Hertz could compensate for the lower energy level. The discrimination scores for the various conditions of bandwidth and delays centered at 250 Hertz are closer in magnitude to the values for the next higher bandwidth centered at 500 Hertz than to the corresponding values of bandwidth centered at 500 Hertz.

Interpretation of the data as plotted as a function of center frequency is not straight forward. The changing locus of the curves across rows in Figures 22, 25 and 28, which represent parametric increases in delay, is presumably due to the effect of delay alone. The presumption is not supported, however, because notwithstanding a nebulous, but general similarity of curve shape across rows, especially those representing bandwidths less than 800 Hertz, there appears to be interesting variations in the shape of the curves. For example, the 400 Hertz bandwidth row for the vowel in Figure 22 appears to reverse its trend as delay increases from 4 to 8 milliseconds. Significant changes across rows may be observed also in Figures 25 and 28 for the plosive and fricative, respectively, but none as significant as in the case of the vowel. How the extent of delay can effect seemingly unsystematic variations in discriminability is not clear.

The changing locus of curves down the columns, representing parametric increases in bandwidth, is presumably due to the effect

of engaging a greater and greater portion of the phoneme spectra in the band of delayed frequency components. The curves appear similar in general down each column with the exception of the curves at the 800 Hertz bandwidth condition. Corresponding columns are also similar for the different phonemes. Corresponding columns are more similar between the plosive and fricative than between either the plosive or fricative and the vowel. The spectra of the plosive and fricative is more similar. The similarity across phonemes suggests that the discrimination scores are partially independent of the phoneme type. The dissimilarities are due to differences in the spectra. The discriminability data trend for the vowel at the 800 Hertz bandwidth phase condition appears to be the reverse of that for other bandwidth conditions. The reversal does not appear for the plosive and fricative. The reversal in the case of the vowel appears to be due to a peak occurring a 1000 Hertz center frequency. The peaking is probably due to the engagement of a vowel formant when bandwidth is increased to 800 Hertz. No other immediate correlation between discriminability and spectra as illustrated in Figure 29 has been observed.

In summary, discriminability has been shown to be a function of delay and bandwidth for all three classes of phonemes investigated. The rank ordering of the phonemes in decreasing order of sensitivity to dispersion is plosive, vowel and fricative. For narrow bandwidths the plosive requires 4 to 8, the vowel 8 to 16 and the fricative 16 to 32 milliseconds of delay for 0.5 discriminability.

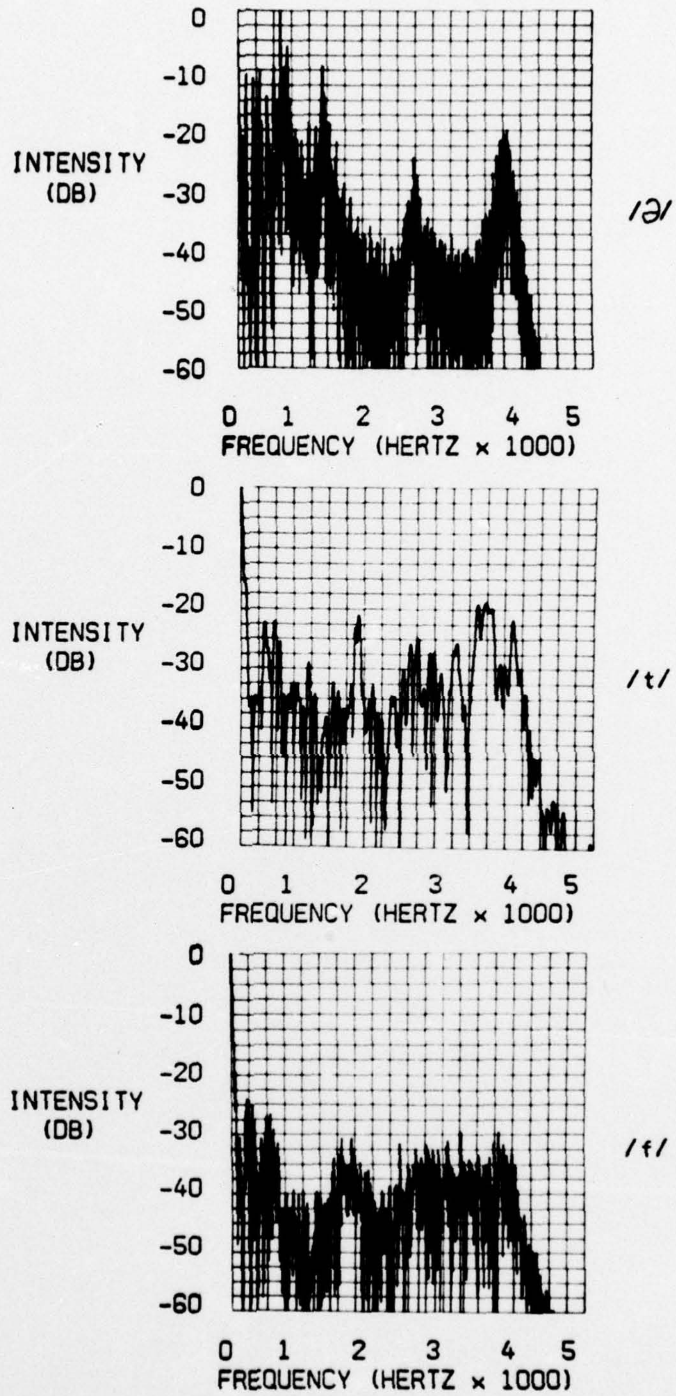


Figure 29. Magnitude spectra of the phonemes used in the study.

For wider bandwidths the respective ranges are 2 to 4, 2 to 4 and 4 to 8 milliseconds. For shorter delays the plosive requires 200 to 400, the vowel 200 to 800 and the fricative 400 to 800 Hertz bandwidth for 0.5 discriminability. For longer delays, the required bandwidth ranges are 100 to 200, 100 to 400 and 100 to 400 Hertz respectively. The difference in range for different bandwidths or delays is probably due to the respective contribution of increased delay or bandwidth on discriminability.

The effect of quantization noise became more apparent with phonemic stimuli. Such noise arises when the samples of a continuous signal are quantized to one of a set of discrete values for storage in computer memory. The quantization noise quality was that of an additive tone. The noise was absent in the standard stimuli. One could argue that the test discriminations would be invalid inasmuch as a subject could discriminate a difference between the two stimuli of the stimulus pair simply by the presence of quantization noise. The author, serving as his own subject however, was aware that the first stimulus of each test pair was the standard stimulus and that the stimuli of all the test pairs were different. The quantization noise was clearly evident as such in most of the stimulus pairs and could be consciously ignored with relatively little effort. Strong evidence to support the validity of the test scores is the internal consistency of the test scores and the external consistency of the test scores with those for words and sentences as will be discussed. Further support of the results is given by their general agreement with the results of informal

discrimination tests in which a slightly phase modified phoneme with consequent quantization noise was used as the standard stimuli. A possible approach to avoiding the quantization noise problem is to add a masking noise to all the stimuli used in the experiment.

Sensitivity to Dispersion in Words

The perceptual quality of words in which a band of frequency components is delayed was observed through informal experiments to be a function of delay for a given center frequency and bandwidth. The change in sound quality was observed for shorter delays at the beginning of words while the change in quality was observed for longer delays over the total duration of the word.

One would expect that the onset of dispersed words beginning with sibilants would be less discriminable than words beginning with plosives because of the noise-like quality of sibilants. The author served as his own subject in an experiment to test the hypothesis. The experiment consisted of two test series, one for the word "DO" and one for the word "SEE." The parameters of the phase distortion were center frequency of 500 Hertz, bandwidth of 400 Hertz and time delays of 8, 16, 32 and 64 milliseconds.

Each test series consisted of four tests, one for each of the four phase conditions. The order of the test presentation within a series was randomized. Each test consisted of 40 randomized pair presentation in which each stimulus of the pair could be either the standard word or the phase modified variant of the word as already discussed. The stimulus pairs were monaurally presented to the right ear at a "comfortable" listening level. The

results of the experiment are illustrated in Figure 30. The plosive onset may be seen from the results of the experiment to be more discriminable than the fricative onset which supports the hypothesis. The delays required for 0.75 discriminability for the two initial sounds are similar to those found in the phoneme experiment.

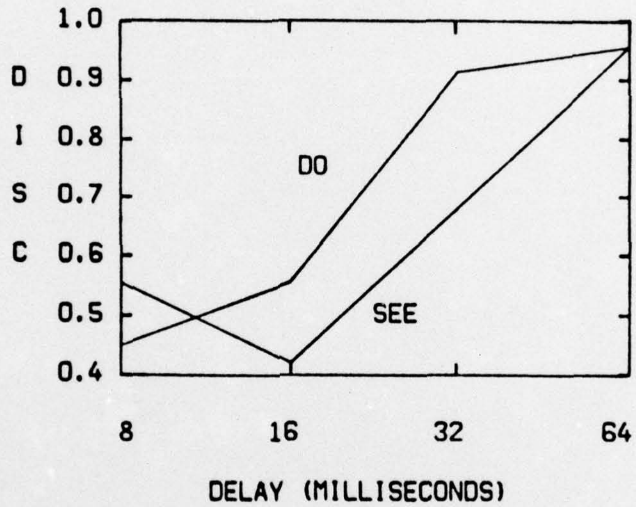


Figure 30. Discriminability of Subject WG to phase modified words as a function of delay.

Sensitivity to Dispersion in a Sentence

An experiment was conducted to determine the discriminability of dispersion in the test sentence, "Men strive but seldom get rich." The three female subjects introduced in Chapter VI served as subjects. A single test in which five individual phase conditions were randomly tested was administered twice to each of the subjects to obtain two discriminability estimates for each phase condition. The subjects were required during the course of the test

to determine whether or not the stimuli of a stimulus pair were the "same" or "different." The phase modification were effected by delaying a 400 Hertz band of frequency components centered at 500 Hertz for different values of delay which included 2, 4, 8 and 16 milliseconds. The stimulus pairs were presented monaurally to the subject's right ear at a level approximating "conversational level" speech. Each phase condition was presented 10 times during the course of the test such that each test consisted of 50 stimulus pair presentations. Half of the test pairs for each phase condition consisted of the standard sentence followed by its phase modified variant whereas the other half consisted of the variant followed by the standard sentence. The stimuli of every pair was consequently different but the subjects were instructed that half of the pairs were the same. This expedient was adopted to reduce the excessive length of the testing period. The results of the experiment are illustrated in Figure 31.

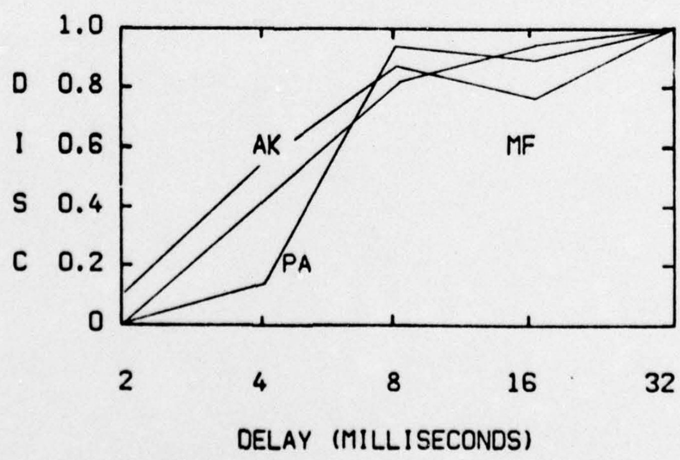


Figure 31. Discriminability of subjects AK, MF and PA to phase modified sentence as a function of delay.

Discriminability is seen to reach the 0.5 level for delays in which a band of frequency components, centered at 500 Hertz and of 400 Hertz bandwidth, is delayed between 4 and 8 milliseconds. The score is similar to that of the vowel under the same phase conditions discussed in the phoneme experiment. The subjects apparently cue on vowel dispersions when discriminating dispersion in continuous speech. Greater dispersions are apparently tolerated in continuous speech in which no absolute standard is available for comparison. The certainty of this conclusion, however, is in doubt because of the lack of control for intensity level between this experiment and that for the phonemes inasmuch as discriminability has been shown to be a function of intensity. Absolute judgments about the quality of the vowel sounds may be the deciding factor in whether or not a given phase dispersion is perceived.

CHAPTER VIII

MONAURAL SENSITIVITY TO DISPERSION IN IMPULSES AND SPEECH: SUMMARY AND CONCLUSIONS

A study was conducted to determine the limits of human phase sensitivity. In review, the role of psychophysics in the design of speech processing systems and the use of computer techniques in psychophysical research were discussed. The history of the "phase sensitivity question" was reviewed and its nature clarified. In part, Ohm postulated the rule by which the ear analyzes sound and Helmholtz added the phase rule as the result of his own work. The validity of Helmholtz' phase rule was discussed and shown to be correct under the conditions he imposed. Helmholtz postulated that the ear would be sensitive to phase under certain conditions.

The work of investigators from the time of Helmholtz to date was reviewed and discussed. The majority of the work supports the fact that the ear is sensitive to phase. Early failures in demonstrating phase sensitivity were probably due to such factors as equipment limitations and theoretical bias. A general intent of much of the early work was the demonstration of various "phase effects." The relevance of such work in the design of speech processing systems was reviewed and the author's philosophical approach to the phase sensitivity question was introduced.

Part of the author's contribution was the philosophical approach to the determination of the limits of phase discriminability in phase distorted speech as a practical goal of phase sensitivity research. The application of computer techniques, including discrete Fourier transforms, to signal processing was discussed. A theoretical analysis of phase distortion was undertaken to provide insight into the nature of phase such that the study could be conducted with a solid understanding of the phenomenon. Two types of experiments were described. The first was concerned with discriminability of dispersion in impulses and the second with dispersion in speech signals. The results of the experiments were discussed in their respective chapters.

The results of the study cannot be compared directly with those of other studies because of basic differences in the stimuli and the type of phase modification used. The study was not designed to explore a new "phase effect" but to determine the limits of phase discriminability under certain conditions as functions of the parameters of an idealized general type of phase distortion. The general exploratory nature of the study preclude extensive theoretical insight into the exact mechanism by which phase discriminability is effected. No general predictions are stated based on the results of the study because of its generally exploratory nature.

A number of criticisms concerning the study should be discussed. The exact characteristics of the Koss PRO-4A headphones used throughout the study were unknown. It was felt that the lack

of information concerning the headphones, however, was a minor problem inasmuch as the testing paradigm required a discrimination between two stimuli both of which were presented with the same headphones. Smaller bandwidths should have been used in the impulse study. Better techniques need to be developed to analyze and display data of three or more independent variables. The theoretical implications of the data obtained during the course of this study have not been exhausted fully simply because of the complexity of the analysis. Better computer graphing techniques should be developed such that the investigator may quickly display the data in various aspects of the independent variables. Statistical reduction may be incorporated profitably in the graphics software.

A possibly interesting variation of the phoneme experiment would be to cepstrually smooth the magnitude of the phoneme spectra so as to remove any confounding effects of rapidly varying magnitude when investigating the effect of center frequency on discriminability. The investigator should have normal hearing such that he may adequately investigate phase discriminability using himself as a subject before conducting extensive tests with other subjects so as to best utilize computer time. Further research is recommended in this area so as to replicate the original study and to further investigate its ramifications.

The intent of the author has been to supply enough general information to permit independent replication of the study. In review, an investigator must obtain a digital recording of the

stimulus to be tested; determine a set of values for the parameters of the phase modifications to be tested; generate the filters necessary to effect the phase modifications; generate a set of phase modified stimuli by convolving the standard stimulus with each of the filters; and finally presenting the phase modified stimuli to a listener under a given testing paradigm. All of the software required to effect the above has been written by the author as an interactive computer program. The program, GSYS.SAI, and its documentation has been archived with the SIPG software at the University of Utah. Independent duplication of the software would require considerable time and expense. Salient features of the software are discussed in Appendix D.

APPENDIX A

ON THE DIFFERENCE IN POPULAR OPINION AS TO WHAT CONSTITUTES OHM'S LAW OF ACOUSTICS

In reviewing the literature concerning Ohm's law of acoustics, it becomes apparent that there may be differences in opinion as to what constitutes the law. It is generally accepted that the law pertains to the frequency analyzing ability of the ear. Confusion arises as to whether or not Ohm (1843) included statements pertaining to the insensitivity of the ear to phase relations among the harmonics of a complex sound, or if such statements were later added by Helmholtz (1863).

To obtain an estimate of the extent of the confusion among authors of acoustical literature, a search by index reference to Ohm's law of acoustics was made of pertinent volumes held at the University of Utah main library. The volumes searched were classified under Sound (534) and Physiological Audition (612.85) within the Dewey Decimal System, and Sound (QC 221-246), Physiological Acoustics (QP 461-471), and Literature of Music - Acoustics and Physics (ML 3805-3817) within the Library of Congress System. An estimate by shelf list of the number of volumes comprising these sections, made by the librarian at the current author's request, was approximately 150 volumes. Of these volumes, 146 were actually searched. Sixteen volumes, or approximately 10

percent, were found to contain the index reference to Ohm's law of acoustics.

The 16 volumes represented a range in publication dates from 1877 to 1970 and represented the writings of 21 authors. The statements of these authors concerning Ohm's law are tabulated in Table 1.

It may be observed from Table 1 that all authors made statements concerning the frequency analyzing ability of the ear in reference to Ohm's law; there is no disagreement among authors on this point. The general interpretation may be exemplified by a statement by Miller (1916):

> The law states: all musical tones are periodic; the human ear perceives pendular vibrations alone as simple tones; all varieties of tone quality are due to particular combinations of a larger or smaller number of simple tones; every motion of the air which corresponds to a complex musical tone or to a composite mass of musical tones is capable of being analyzed into a sum of simple pendular vibrations, and to each simple vibration corresponds a simple tone which the ear may hear.

It is generally accepted that the type of analysis performed by the ear is a Fourier analysis.

Four authors made statements to the effect that Ohm's law also pertains to the insensitivity of the ear to phase. Bekesy (1960), citing Helmholtz and Ohm, stated that "According to Ohm's law, the perception of a complex sound depends upon its analysis into single tones and is entirely independent of the phase relations of these components . . . ." Bacus (1969) made a similar statement in reference to Ohm's law in citing Bekesy. Rayleigh (1877) stated in connection with his reference to Ohm's law that ". . . , within

the limits of audibility the relative phases of the various components would be a matter of indifference." Stevens and Davis (1938) wrote that the usual assertion of Ohm's law is that "the ear tends to analyze the components of a complex sound regardless of their phase-relations."

In contrast to the statements of the four authors quoted above, four other authors implied that the phase sensitivity statements were later added by Helmholtz. Dittrich (1963) wrote that Helmholtz, as a result of his experiments with synthetic sounds, determined that "contrary to what was believed up to this time, the musical timbre of a sound depends on the number and the intensity of its harmonics and not on their phase relation, ..." Richardson (1927) also implied that phase statements were not part of Ohm's law by writing that "If Ohm's law be true, the question is naturally asked, what influence have the respective phases of the components of the note on the impression of its quality?" It would appear from the statement that the effect of phase on quality was not part of Ohm's law. In support of this conclusion, Richardson is further cited as having written that "Helmholtz and also Konig attacked this problem experimentally."

Wever (1949) wrote that Helmholtz supported his resonance theory of hearing

> . . . from his experiments on the synthesis of vowels and other complex sounds. He showed that such sounds could be imitated by exciting simultaneously a number of tuning forks suitably chosen in frequency and intensity. Only frequency and intensity were significant variables here; the phase relations could be changed at will without noticeable alteration of what he referred to as the musical quality. This ineffectiveness of phase, in Helmholtz's view, was the result of the fact that the ear analyzes the sound into its

component simple tones, and phase relations are lost in the process.

The last author cited to imply that Helmholtz added statements concerning the sensitivity of the ear to phase was A. Wood (1940), who stated that "Helmholtz satisfied himself that relative phase is without effect on the quality of a note."

Two authors definitely stated that the additions were made by Helmholtz. Lamb (1910) wrote "According to Helmholtz the influence of phase is inappreciable." Miller (1916) wrote that "As the result of elaborate investigation, Helmholtz added the following law: the quality of a musical tone depends solely on the number and relative strength of its partial simple tones, and in no respect on their differences of phase."

In support of the statements made by the four authors just cited, A. B. Wood (1930) categorically stated that "Ohm's law says nothing about the relative phases of the harmonic components into which the ear analyses the sound, ..." This statement is in direct conflict with those of Bacus, Bekesy, Rayleigh, and Stevens and Davis.

The interpretations of the eight authors who made no reference to the ear's insensitivity to phase in connection with their statements of Ohm's law of acoustics is not clear. It is not unreasonable to think, however, that the lack of reference implies that their interpretation pertains only to the frequency analyzing ability of the ear.

In summary, 19 authors (or collaborators) were cited who made reference to Ohm's acoustical law. Of these authors 58 percent

made statements concerning the sensitivity of the ear to phase in reference to the law. Of this group 36 percent stated that Ohm's law includes phase statements, while the other 64 percent stated, or implied, that the law does not include phase statements, but that the statements were later added by Helmholtz. The interpretations of the authors who made no phase statements is not clear.

Pertinent journal literature was also searched to further ascertain the extent of the confusion as to what constitutes Ohm's law of acoustics. Completeness of the journal survey is supported by cross-referencing among the articles. Five articles are cited which state that Helmholtz added the statements concerning phase indifference of the ear. No journal articles were found that stated the contrary.

Beasley (1931) understood that the effect of phase upon tone quality was of interest even before Helmholtz' research but wrote that Helmholtz "announced the much-quoted law: 'the quality of a musical tone depends on the number and relative strength of its partial simple tones, and not on their differences of phase.'" Beasley referred to this statement as Helmholtz' "law."

Chapin and Firestone (1934) wrote that "since the work of Helmholtz it has been widely accepted that the tone quality of a complex tone is independent of the relative phases of its harmonics." This statement implies that the concept originated with Helmholtz.

Schouten (1939) referred to the frequency analyzing ability of the ear as Ohm's law and further added that "According to a rule

proposed by Helmholtz the sound perception will further depend entirely on the relative intensity with which the different components occur and will be independent of the relative phase of these components."

Mathes and Miller (1947) stated the following in reference to Helmholtz' additions concerning the sensitivity of the ear to phase:

> Ohm's acoustic law states that the ear analyzes a complex sound into simple tones independent of each other. Helmholtz as the result of carefully conducted experiments, amplified this statement with the declaration that ". . . the quality of the musical portion of a compound tone depends solely on the number and relative strength of its partial simple tones, and in no respect on their differences of phase."

Craig and Jeffress (1962) implied Helmholtz' additions pertaining to phase by writing that Helmholtz' theory did not account for monaural phase effects and that "he avoided the difficulty by denying their existence." These authors cited Helmholtz as having "classified an admittedly heard phase determined auditory effect (the faint beats produced by a tone and its mistuned octave) as only an apparent exception to his phase rule."

In summary, all relevant journal articles, checked for inclusiveness by internal cross-referencing, indicated unanimously that Ohm's law of acoustics does pertain to the frequency analyzing ability of the ear and that Helmholtz later added the statements concerning the phase indifference of the ear as a result of his own experimental work.

A careful reading of Ohm and Helmholtz reveals that Ohm did not in fact mention the sensitivity of the ear to phase but that

such statements were later added by Helmholtz as a result of his own work.

In conclusion, it appears that the vast majority of authors from the time of Helmholtz to date have held the interpretation of Ohm's law summarized immediately above. There appear to be a few authors, however, who hold different interpretations such as Bekesy, Rayleigh, and Stevens and Davis. Whether or not these authors hold that the phrase "Ohm's law of acoustics" includes the original statements of Ohm plus later additions by Helmholtz is not clear. It is difficult to believe that these authors, in particular, could make an error of this nature. It seems more likely that they are using the phrase "Ohm's law of acoustics" to mean in a very general sense the combined contributions of Ohm and Helmholtz. There is precedence in such use of terms. For example, Newton's "universal law of gravity" is correct only within certain limits. Later contributions by Einstein have resulted in a more correct "universal law of gravity," yet the phrase "universal law of gravity" is applied with equal correctness to the concepts of Newton, the concepts of Einstein, or the concepts of both combined.

It would appear then that in referring to Ohm's law of acoustics the investigator is at liberty to apply the phrase "Ohm's law of acoustics" in a more restricted, generally used, and correct sense in reference to the statements of Ohm, or in a looser sense in reference to the contributions of both Ohm and Helmholtz. The investigator will be in good company with either use of the phrase as this paper has demonstrated.

TABLE 1

STATEMENTS OF SELECTED AUTHORS IN REFERENCE TO OHM'S LAW OF ACOUSTICS

Author	Reference Type and Page					
	1	2	3	4	5	6
Bacus (1969)	99	99				
Bekesy (1960)	471	471				
Denes (1963)	108					X
Dittrich (1963)	32		33			
Lamb (1910)	2			4		
Miller (1916)	62			62		
Nordmark (1970)	58					X
Rayleigh (1877)	443	443				
Richardson (1927)	232		233			
Small (1970)	4					X
Stephens and Bates (1950)	7					X
Stevens and Davis (1938)	20	203				
Tonedorf (1970)	232					X
Ward (1970)	438					X
Wever (1949)	26		28			
Wever and Lawrence (1954)	146					X
Whitfield (1967)	143					X
Wood, A. B. (1930)	356				357	
Wood, A. (1940)	340		364			

1 Statements that Ohm's law pertains to aural frequency analysis.

2 Statements that Ohm's law also pertains to the ear's insensitivity to phase.

3 Statements that imply that Helmholtz added phase statements.

4 Statements that Helmholtz added phase statements.

5 Statements that Ohm's law does not include any reference to the sensitivity of the ear to phase.

6 No reference to phase in reference to Ohm's law.

APPENDIX B

AN INTRODUCTION TO DIGITAL COMPUTERS

A digital computer consists basically of a memory, a processing unit, and associated peripheral devices. The memory consists of a block of addressable slots, called words, in which numbers may be stored or retrieved. The processing unit acts upon the numbers in memory to perform a logical or arithmetical operation and to return the result back to memory. The action of the processing unit is controlled by a sequential list of instructions, called a program, which is also stored in memory. The distinctive feature of a computer that makes it more than a programmed calculator and endows it with its computational power is that the program may modify itself under program control.

The computational power of computers is related to the number or instructions in their repertoire, the speed at which they may execute the instructions, and their memory size. Modern computers are able to execute instructions at rates approaching one-million per second, and have memory sizes in the order of a quarter-million words. Memory size is colloquially specified in "K," where one-K represents 1024 words, e.g., 96 K of memory.

Computers use a binary (base-2) number system to internally represent numbers because of the ease in which such numbers may be electronically represented. Only two digits are allowed in the

binary number system, 0 and 1. The digits may be represented, for example, by the state of a bistable electrical device such as a switch. A ten state switch would be required to represent a digit in the base-10 number system. Binary digits are colloquially referred to as bits. In any given computer, each memory word contains a given number of bits. The greater the number of bits, the larger the number that can be stored.

Computers have a large variety of associated peripheral devices. Many of the devices are for input-output such as teletypes, high-speed line printers, oscilloscopes, analog-to-digital converters, digital-to-analog converters, etc. Other peripherals include magnetic tape and disk mass storage devices. The data base for a given computer problem is typically greater then what can be stored at one time in memory. The solution to the problem is to store the data base on disk, sequentially read into computer memory only as much of the data that can be handled, process the data, and then sequentially write the data back out to disk.

Computer applications appear to be limited only by the complexity of the program which a programmer may devise, its speed and memory size, and the characteristics of its input-output system.

APPENDIX C

DATA TABLES

TABLE 2

DISCRIMINABILITY OF SUBJECTS AK, MF AND PA BETWEEN AN IMPULSE AND IMPULSES IN WHICH A BAND OF FREQUENCY COMPONENTS ARE DELAYED AS FUNCTIONS OF BANDWIDTH AND DELAY.

BANDWIDTH (HERTZ)	DELAY (MILLISECONDS)			
	0.0625	0.125	0.25	0.5
SUBJECT: AK				
100	0.47	0.64	0.63	0.85
200	0.56	0.66	0.67	0.91
400	0.55	0.69	0.84	0.91
800	0.50	0.58	0.82	0.94
SUBJECT: MF				
100	0.50	0.49	0.65	0.76
200	0.52	0.71	0.68	0.85
400	0.48	0.53	0.68	0.90
800	0.48	0.61	0.79	0.73
SUBJECT: PA				
100	0.53	0.53	0.60	0.93
200	0.52	0.46	0.69	0.84
400	0.54	0.50	0.54	0.98
800	0.52	0.52	0.94	0.97

TABLE 3

DISCRIMINABILITY OF SUBJECTS AK, MF AND PA BETWEEN AN IMPULSE AND A PHASE MODIFIED IMPULSE AS A FUNCTION OF INTENSITY. THE PARAMETERS ARE 500 HERTZ CENTER FREQUENCY, 400 HERTZ BANDWIDTH AND 0.25 MILLISECONDS DELAY.

	RELATIVE INTENSITY (DB)		
SUBJECT	-6	0	6
AK	0.55	0.67	0.87
MF	0.49	0.68	0.72
PA	0.48	0.49	0.97

TABLE 4

DISCRIMINABILITY OF SUBJECT WG BETWEEN AN IMPULSE AND PHASE MODIFIED IMPULSES AS A FUNCTION OF INTENSITY. PHASE IS MODIFIED AS FUNCTIONS OF BANDWIDTH AND DELAY.

DELAY (mSec)	INTENSITY (dB SL)					
	68	71	74	77	80	83
BW = 100 HERTZ						
0.0625				0.550	0.725	0.783
0.125		0.500	0.617	0.700	0.725	0.850
0.25	0.550	0.650	0.817	0.983	0.908	0.983
0.5	0.650	0.917	0.967	1.000		
BW = 200 HERTZ						
0.0625			0.467	0.550	0.750	0.792
0.125	0.517	0.567	0.671	0.839	0.922	0.894
0.25	0.567	0.583	0.833	0.992	0.983	1.000
0.5	0.917	0.933	0.983	1.000		
BW = 400 HERTZ						
0.0625	0.467	0.500	0.517	0.625	0.750	0.833
0.125	0.508	0.489	0.794	0.850	0.972	
0.25	0.561	0.817	0.894	0.933	0.961	
0.5	0.917	1.000	0.992			
BW = 800 HERTZ						
0.0625			0.467	0.517	0.633	0.783
0.125	0.483	0.550	0.600	0.750	0.867	0.850
0.25	0.467	0.683	0.950	0.950	1.000	
0.5		1.000		1.000		

TABLE 5

DISCRIMINABILITY OF SUBJECT WG TO DISPERSION IN IMPULSES AS A FUNCTION OF BANDWIDTH AND DELAY FOR A FIXED CENTER FREQUENCY OF 500 HERTZ AND SENSATION LEVEL OF 80 DB USING A CLASSICAL CONSTANT STIMULI EXPERIMENTAL DESIGN.

DELAY (mSEC)	BANDWIDTH (HERTZ)			
	100	200	400	800
0.0625	0.51	0.55	0.58	0.68
0.125	0.52	0.61	0.79	0.92
0.25	0.76	0.92	0.95	0.96
0.5	0.89	0.98	1.00	1.00

TABLE 6

MEAN AND STANDARD DEVIATION OF THE DISCRIMINABILITY OF SUBJECT WG BETWEEN AN IMPULSE AND IMPULSES IN WHICH A BAND OF FREQUENCY COMPONENTS ARE DELAYED AS FUNCTIONS OF CENTER FREQUENCY, BANDWIDTH AND DELAY.

DELAY (mSEC)	BANDWIDTH (HERTZ) 50	100	200	400	800	1600	3200
FC = 250 HERTZ							
0.0625	0.4833	0.4667	0.5500	0.6000			
	0.0204	0.0540	0.0354	0.0707			
0.125	0.5250	0.5417	0.6167	0.7333			
	0.0354	0.0816	0.1633	0.1080			
0.25	0.7833	0.7417	0.8917	0.9500			
	0.0816	0.1744	0.2041	0.0612			
0.5	0.9000	0.9583	0.9583	0.9917			
	0.0354	0.0204	0.0408	0.0204			
FC = 500 HERTZ							
0.0625	0.4833	0.6500	0.6500	0.7083	0.6278		
	0.0204	0.1841	0.1225	0.1124	0.0068		
0.125	0.8500	0.7333	0.9222	0.9722	0.9056		
	0.2121	0.0408	0.0360	0.0136	0.0593		
0.25	1.0000	0.9222	0.9722	0.9611	0.9917		
	0.0000	0.0491	0.0272	0.0758	0.0204		
0.5	1.0000	1.0000	1.0000	1.0000	1.0000		
	0.0000	0.0000	0.0000	0.0000	0.0000		
FC = 1000 HERTZ							
0.0625	0.5750	0.8417	0.7250	0.8083	0.7750	0.9125	
	0.0354	0.2131	0.1275	0.1242	0.0354	0.0530	
0.125	0.9083	0.8583	0.8750	0.9917	0.9500	0.9875	
	0.0540	0.1671	0.1061	0.0204	0.0707	0.0177	
0.25	1.0000	0.9917	0.9917	1.0000	1.0000	1.0000	
	0.0000	0.0204	0.0204	0.0000	0.0000	0.0000	
0.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

TABLE 6 (CONTINUED)

DELAY (mSEC)	BANDWIDTH (HERTZ) 50	100	200	400	800	1600	3200
FC = 2000 HERTZ							
0.0625	0.6063	0.7100	0.7375	0.9063	0.8625	0.8813	0.7250
	0.1293	0.2706	0.1920	0.1083	0.1031	0.1139	0.1768
0.125	1.0000	0.8563	0.9417	0.9500	0.9667	0.9750	0.9167
	0.0000	0.1672	0.0540	0.1458	0.0408	0.0354	0.0408
0.25	1.0000	0.9833	1.0000	1.0000	1.0000	1.0000	1.0000
	0.0000	0.0204	0.0000	0.0000	0.0000	0.0000	0.0000
0.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

TABLE 7

DISCRIMINABILITY OF SUBJECT WG BETWEEN A STANDARD NEUTRAL VOWEL STIMULUS AND PHASE MODIFIED VARIANTS OF THE THE VOWEL IN WHICH A BAND OF FREQUENCY COMPONENTS ARE DELAYED AS FUNCTIONS OF CENTER FREQUENCY, BANDWIDTH, AND DELAY.

BANDWIDTH (HERTZ)	DELAY (MILLISECONDS) 2	4	8	16
FC = 250 HERTZ				
50	0.000	0.000	0.000	0.000
100	0.000	0.000	0.067	0.333
200	0.000	0.067	0.167	0.833
400	0.000	0.400	0.800	0.933
FC = 500 HERTZ				
100	0.000	0.000	0.167	0.200
200	0.033	0.033	0.333	0.600
400	0.000	0.567	0.267	0.700
800	0.200	0.800	0.467	0.533
FC = 1000 HERTZ				
100	0.000	0.033	0.067	0.050
200	0.000	0.033	0.300	0.400
400	0.033	0.267	0.333	0.700
800	0.333	0.967	1.000	1.000
FC = 2000 HERTZ				
100	0.000	0.033	0.067	0.133
200	0.067	0.033	0.100	0.400
400	0.067	0.167	0.633	0.767
800	0.067	0.467	0.900	0.967

TABLE 8

DISCRIMINABILITY OF SUBJECT WG BETWEEN A STANDARD PLOSIVE, /t/, AND PHASE MODIFIED VARIANTS OF THE PLOSIVE IN WHICH A BAND OF FREQUENCY COMPONENTS ARE DELAYED AS FUNCTIONS OF CENTER FREQUENCY, BANDWIDTH, AND DELAY.

BANDWIDTH (HERTZ)	DELAY (MILLISECONDS) 2	4	8	16
FC = 250 HERTZ				
50	0.050	0.000	0.000	0.000
100	0.000	0.000	0.050	0.050
200	0.100	0.500	0.950	1.000
400	0.100	0.500	0.850	1.000
FC = 500 HERTZ				
100	0.050	0.150	0.950	1.000
200	0.050	0.100	0.900	1.000
400	0.400	0.800	1.000	1.000
800	0.600	0.900	1.000	1.000
FC = 1000 HERTZ				
100	0.100	0.050	0.100	0.300
200	0.050	0.150	0.250	0.850
400	0.150	0.000	0.400	0.950
800	0.200	0.750	0.900	1.000
FC = 2000 HERTZ				
100	0.050	0.067	0.100	0.100
200	0.150	0.133	0.050	0.050
400	0.050	0.067	0.350	0.950
800	0.300	0.500	1.000	1.000

TABLE 9

DISCRIMINABILITY OF SUBJECT WG BETWEEN A STANDARD FRICATIVE, /f/, AND PHASE MODIFIED VARIANTS OF THE FRICATIVE IN WHICH A BAND OF FREQUENCY COMPONENTS ARE DELAYED AS FUNCTIONS OF CENTER FREQUENCY, BANDWIDTH, AND DELAY.

BANDWIDTH (HERTZ)	DELAY (MILLISECONDS) 4	8	16	32
FC = 250 HERTZ				
50	0.000	0.000	0.000	0.000
100	0.100	0.150	0.100	0.850
200	0.300	0.100	0.850	1.000
400	0.050	0.200	0.900	1.000
FC = 500 HERTZ				
100	0.100	0.000	0.300	1.000
200	0.050	0.050	0.300	1.000
400	0.400	0.250	0.850	1.000
800	0.550	0.650	1.000	1.000
FC = 1000 HERTZ				
100	0.000	0.000	0.000	0.600
200	0.033	0.000	0.550	1.000
400	0.000	0.167	1.000	1.000
800	0.033	0.167	0.950	1.000
FC = 2000 HERTZ				
100	0.000	0.000	0.050	0.050
200	0.067	0.000	0.150	1.000
400	0.067	0.200	0.750	1.000
800	0.233	0.650	0.850	1.000

APPENDIX D

SELECTED SOFTWARE DESCRIPTION

GSYS

GSYS is an interactive computer program used to process, replay, and display audio signals, and to administer psychophysical tests. The program is written in SAIL, an ALGOL-like high-level computer language. The program types a herald on the controlling terminal to indicate its readiness to accept a command. After executing any given command, the program again types the herald and re-enters its waiting state. A list of the available commands may be obtained by typing a question-mark. Typing the first character of a command results in the automatic typing of the rest of the command name. The program will request parameter value input if required by the given command. The program is organized around a large program buffer which is used to store audio data for processing, replay, etc. The audio data are organized as audio files for manipulation. Provisions are made for the absolute accessing of audio data on disk. A description of the audio file system follows under the title of FILES. System commands include the following:

AVGOUT — Remove any DC bias from the audio data stored in the program buffer.

CONVOLVE — Convolve a given source file with a given kernel file. The results are left in the program buffer.

DRIVE	Change the default drive number.
EXPUNGE	Delete an audio file stored on disk.
FREQUENCY	Set the frequency of the digital-to-analog converter.
GET	Copy an audio file from disk to the program buffer.
INPUT	Copy audio data from disk to the program buffer using absolute disk addressing.
KERGEN	Generate the frequency domain representation of the idealized filters used in the study.
LIST	List the disk file directory.
MAKE	Generate special audio signals in the program buffer.
NWORDS	Specify the number of words of data in the program buffer in reference to the buffer's lower index to be used in processing, replay, etc.
ONOFF	Window the audio data in the program buffer with a trapezoidal window.
PLOT	Display the audio data in the program buffer on the storage oscilloscope. See the documentation for PLOT in GMAC.
RMS	Type the RMS value of the data in the program buffer.
SAVE	Copy the program buffer to a disk file.
TEST	Psychophysical test. See the documentation for IAFCT in GMAC.
ZERO	Clear the program buffer.
1IDX	Set the program buffer's lower index.
2IDX	Set the program buffer's upper index.
5T'S	Design a filter using the 5T's method. The KERGEN command is first called to specify the idealized filter characteristics.
$PLAY	Play the data in the program buffer through the digital-to-analog converter.

GMAC

GMAC is the author's Assembly language subroutine library. The subroutines are used by GSYS. The following subroutines are selected for description:

SUBROUTINE IAFCT(DRIVE,X,RLST)

IAFCT is an interactive audio file comparison test subroutine. It presents a number of audio file-pairs for discrimination testing through DA16A. The following commands are provided:

BEGIN TEST	Start the test program.
DELAY	Set the time delay between signal pairs and between signals of the pair.
INDEX	Set the beginning index into the signal-pair randomizer.
LIST RESULTS	Type a list of the signal-pairs presented, the number of times each pair-type was presented, and the number of correct discriminations for each pair-type.
NUMBER	Set the number of times that the signal-pair list is randomized and presented.
REF TABLE	Input a list of signal-pairs from the terminal to be presented in the test. Terminate the list with -1.
TTY MODE	TRUE (-1): Get listener's response from the TTY for each signal-pair presentation. FALSE (0): Present the whole test without TTY interaction.
?	Type the command list.

SUBROUTINE APF1(X,Y,PWR2,FC,BW,TD,SFREQ)

APF1 specifies the frequency domain characteristics of an ideal filter designed to give a constant time delay (TD) to a band of frequency components (BW) centered around a given center frequency (FC). The magnitude is returned in X, and the phase in Y.

SUBROUTINE CONV(X,PWR2,NREC,RPK,RBLK,WPK,WBLK)

CONV performs linear convolution. It assumes that the kernel resides in the first N words of array X, where N=2PWR2. The subroutine types the record number, the input disk block, the output disk block, and the number of packing overflows for each record.

SUBROUTINE PLOT(N,NWORDS,DELTA)

PLOT is an interactive plotting subroutine with the following commands:

C *Grid complexity.*

E Expand and plot with one-shot settings of grid coordinates.

H Horizontal label.

I Intensity.

M Compute new Y-coordinates for each plot.

O Offset Y-coordinates from average.

P Plot.

T Grid type.

U Multiple plot on same grid image.

V Vertical label.

Y Y-coordinates from the TTY.

? Type this list.

<carriage return> Subroutine return.

FILES

FILES is a simple file system designed to facilitate the storage and retrieval of data on private disk packs. The system formats a given pack by storing a directory image on the disk pack beginning at block 200. The first two words of the directory image serve as the directory header with the following format:

WORD 0: Left justified SIXBIT key-word "AUDIO."

WORD 1: Left justified SIXBIT pack name.

Following the directory header are 1023 2-word file descriptor blocks with the following format:

WORD 0: LH: User supplied data.
RH: File block location.

WORD 1: BITS 0-12: Date of file creation.
BITS 13-35: Length of file (WORDS).

Data is stored on disk beginning at block 1000. Data may be stored up to, and including, block 39999. The files on a given pack are referred to by a file number ranging from 1 to 1023. When a new file is created the system finds and returns the first free block number above the file occupying the highest numbered block.

SUBROUTINE SETDIR(ERROR,DRIVE,NAME) formats a given disk pack for the file system and clears the directory.

SUBROUTINE DELF(ERROR,DRIVE,FILNUM) deletes a file from the directory.

SUBROUTINE DIR(ERROR,DRIVE,MODE) types the directory information on the TTY.

SUBROUTINE ERR(ERROR) types the error message associated with an error number on the TTY. *ERR is a NO-OP if ERROR=0.*

SUBROUTINE OPENR(ERROR,DRIVE,FILNUM,BLOCK,NWORDS,DATA) returns disk addressing information and data pertaining to a given file number. Used to read old data from disk.

SUBROUTINE OPENW(ERROR,DRIVE,FILNUM,BLOCK,NWORDS,DATA,AGE) returns disk addressing information and associated pertinent information for a given file number. Used to write new data on disk.

REFERENCES

Allen, J. B. 1972. Removing the characteristic buzz from synthetic speech. Unpublished Bell Telephone Laboratories technical memorandum.

Bacus, J. 1969. The Acoustical Foundations of Music. New York: W. W. Norton.

Beasley, W. 1930a. The monaural phase effect with pure binary harmonies I. Frequency ratio 2:3. Journal of the Acoustical Society of America. 1, no. 2: 385-402.

Beasley, W. 1930b. The influence of phase on tone quality. Journal of the Acoustical Society of America. 1: 170(A).

Beasley, W. 1931. Differential responses to cyclic phase variations in compound sounds. Journal of General Psychology. 5: 329-349.

von Bekesy, G. 1960. Experiments in Hearing. Translated and edited by E. G. Wever. New York: McGraw-Hill.

Boring, E. G. 1938. Perspective. In Hearing. Authored by S. S. Stevens and H. Davis. New York: John Wiley and Sons.

Callahan, M. W. 1975. Two dimensional processing of acoustic signals. IEEE 1975 Region Six Conference Proceedings.

Chapin, E. K. and Firestone, F. A. 1934. The influence of phase on tone quality and loudness; the interference of subjective harmonics. Journal of the Acoustical Society of America. 5, no. 3: 173-180.

Cole, E. R. 1973. The removal of unknown image blurs by homomorphic filtering. Unpublished doctoral dissertation, University of Utah.

Cooley, J. W. and Tukey, J. W. 1965. An algorithm for the machine calculation of complex Fourier series. Mathematics of Computation. 19: 297-301.

Craig, J. H. 1961. The effect of phase on the quality of a two-component tone. Unpublished doctoral dissertation, University of Texas.

Craig, J. H. and Jeffress, L. A. 1962. Effect of phase on the quality of a two-component tone. Journal of the Acoustical

Society of America. 34, no. 11: 1752-1760.

Denes, P. B. and Pinson, E. N. 1963. The Speech Chain. Bell Telephone Laboratories.

Dittrich, F. L. 1963. Biophysics of the Ear. Springfield: Charles C. Thomas.

Flanagan, J. L. 1950. Effects of phase distortion of speech. Unpublished master's thesis, Massachusetts Institute of Technology.

Flanagan, J. L. 1951. Effect of delay distortion upon the intelligibility and quality of speech. Journal of the Acoustical Society of America. 23, no. 3: 303-307.

Goldstein, J. L. 1967. Auditory spectral filtering and monaural phase perception. Journal of the Acoustical Society of America. 41, no. 2: 458-479.

Guillemin, E. A. 1963. Theory of Linear Physical Systems. New York: John Wiley and Sons, Inc.

Hansen, V. and Madsen, E. R. 1974. On aural phase detection. Journal of the Audio Engineering Society. 22, no. 1: 10-14.

von Helmholtz, H. L. F. 1863. On the Sensations of Tone. The second 1885 English edition conformal to the the fourth German edition of 1877 with material added by Alexander J. Ellis (Editor and translator). New York: Dover Publications, 1954.

Koenig, R. 1881. In A. J. Ellis (Editor and translator) H. v. Helmholtz, On the Sensation of Tone. New York: Dover Publications, 1954.

Lamb, H. 1910. The Dynamical Theory of Sound. London: Edward Arnold.

Lane, C. E. 1930. Phase detection in telephone apparatus. Bell System Technical Journal. 9: 493-521.

Lewis, D. and Larsen, M. J. 1937. The cancellation, reinforcement, and measurement of subjective tones. Proceedings of the National Academy of Science. 23: 415-421.

Licklider, J. C. R. 1951. Basic correlates of the auditory stimulus. Handbook of Experimental Psychology. Edited by S. S. Stevens. New York: John Wiley and Sons, Inc.

Licklider, J. C. R. 1957. Effects of changes in the phase pattern upon the sound of a 16-harmonic tone. Journal of the Acoustical Society of America. 29, no. 6: 780(A).

Mathes, R. C. and Miller, R. L. 1947. Phase effects in monaural perception. Journal of the Acoustical Society of America. 19, no. 5: 780-797.

Miller, D. C. 1916. The Science of Musical Sounds. New York: Macmillan.

Miller, N. J. 1973. Removal of noise from a voice signal by synthesis. Unpublished Doctoral dissertation, University of Utah.

Nordmark, J. O. 1970. Time and frequency analysis. Foundations of Modern Auditory Theory. Edited by J. V. Tobias. New York: Academic Press.

Ohm, G. S. 1843. Uber die Definition des Tones, nebst daran geknupfter Theorie der Sirene und ahnlicker tonbildender Vorrichtungen. Annalen der Physik und Chemie. 59, ser. 2: 513-565.

Oppenheim, A. V.; Schafer, R. W.; and Stockham, T. G., Jr. 1968. Nonlinear filtering of multiplied and convolved signals. Proceedings of the IEEE. 56, no. 8: 1264-1291.

Papoulis, A. 1962. The Fourier Integral and its Applications. New York: McGraw-Hill.

Rabiner, L. R.; Schafer, R. W.; and Rader, C. M. 1969. The chirp z-transform algorithm. IEEE Transactions on Audio and Electroacoustics. AU-17: 86-92.

Rayleigh, Lord. 1877. The Theory of Sound. First American edition, the first edition being published in 1877. New York: Dover Publications, 1945.

Rayleigh, Lord. 1907. On our perception of sound direction. Philosophical Magazine. 6 Ser., 13: 214-232.

Richardson, E. G. 1927. Sound. London: Edward Arnold.

Schouten, J. F. 1939. Synthetic sound. Philips Technical Review. 4, no. 6: 167-173.

Schroeder, M. R. 1959. New results concerning monaural phase sensitivity. Journal of the Acoustical Society of America. 31, no. 11: 1579(A).

Schroeder, M. R. 1966. Vocoders: Analysis and Synthesis of Speech. Bell Telephone System Technical Publications, Monograph 5184.

Small, A. M. 1970. Periodicity pitch. Foundations of Modern Auditory Theory. Edited by J. V. Tobias. New York: Academic

Press.

Steinberg, J. C. 1930. Effects of phase distortion telephone quality. Bell System Technical Journal. 9: 550-566.

Stephens, R. W. B. and Bates, A. E. 1950. Wave Motion and Sound. London: William Clowes and Sons.

Stevens, S. S. and Davis, H. 1938. Hearing. New York: John Wiley and Sons.

Stockham, T. G., Jr. 1966. High-speed convolution and correlation. Spring Joint Computer Conference. AFIPS Proceedings. 28: 229-233.

Stockham, T. G., Jr. 1969. High-speed convolution and correlations with applications to digital filtering. In B. Gold and C. M. Rader, Digital Processing of Signals. New York: McGraw-Hill.

Stockham, T. G., Jr. 1972. Restoration of old acoustic recordings and photographic image deblurring by means of digital signal processing. 1972 IEEE Intercon Digest.

Stockham, T. G., Jr.; Cannon, T. M.; and Ingebretsen, R. B. 1975. Blind deconvolution through digital signal processing. Proceedings of the IEEE. 63, no. 4: 678-692.

Thompson, S. P. 1877. On binaural audition. Philosophical Magazine. 5 Ser., 4: 274-276.

Tonndorf, J. 1970. Cochlear mechanics and hydrodynamics. Foundations of Modern Auditory Theory. Edited by J. V. Tobias. New York: Academic Press.

Trimmer, J. D. and Firestone, F. A. 1937. An investigation of subjective tones by means of the steady tone phase effect. Journal of the Acoustical Society of America. 9: 24-29.

Ward, W. D. 1970. Musical perception. Foundations of Modern Auditory Theory. Edited by J. V. Tobias. New York: Academic Press.

Warnock, R. B. 1973. Audio Console Instruction Manual. Unpublished technical manual, Department of Computer Science, University of Utah.

Wever, E. G. 1949. Theory of Hearing. New York: Dover Publications, 1970.

Wever, E. G. and Lawrence, M. 1954. Physiological Acoustics. Princeton: Princeton University Press.

Whitfield, J. C. 1967. The Auditory Pathway. London: Edward Arnold.

Wightman, F. L. and Green, D. M. 1974. The perception of pitch. American Scientist. 62: 208-215.

Wood, A. B. 1930. A Textbook of Sound. New York: Macmillan.

Wood, A. 1940. Acoustics. London: Blackie and Son Limited.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (14) UTEC-CSc-75-142	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER (9)
4. TITLE (and Subtitle) (6) MONAURAL SENSITIVITY TO DISPERSION IN IMPULSES AND SPEECH.		5. TYPE OF REPORT & PERIOD COVERED Technical Report.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) William Harvey Greet		8. CONTRACT OR GRANT NUMBER(s) (15) DAHC15-73-C-Ø363, F30602-70-C-Ø300
9. PERFORMING ORGANIZATION NAME AND ADDRESS Computer Science Department University of Utah Salt Lake City, Utah 84112		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS ARPA Order #2477
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, Virginia 22209		12. REPORT DATE (11) December 1975
		13. NUMBER OF PAGES 151 (12) 154p.
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

404 949

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Little is known of the ear's phase sensitivity limits outside the well known fact that excessive dispersion in speech signals results in chirp-like sound quality. The limits of phase sensitivity may be determined in reference to the parameters of an idealized dispersion since phase shifts are nothing more than delays in the short-time spectra of a signal. A number of discrimination tests were conducted which contrasted standard stimuli with phase modified variants of the stimuli to determine the limits of phase sensitivity. The idealized dispersion was that of delaying a band of frequency components such

DD FORM 1 JAN 73 1473A EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED (cont on p 1473B)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(cont fr p1473A)

that discriminability could be measured as functions of center frequency, bandwidth and delay. The stimuli used in the tests included an impulse, phonemes, words, and a sentence.

Sensitivity to dispersion in impulses was shown to be dependent on intensity, center frequency, bandwidth, and delay. Discrimination scores half way between chance and perfect performance were achieved for dispersed impulses with delayed bands of frequency components centered between 250 and 500 ~~Hertz~~ Hz and delayed between 0.0625 and 0.125 milliseconds. Sensation level of the stimulus pairs was 80 dB. The effect of bandwidth on discriminability appears to reach a plateau at 100 ~~Hertz~~ Hz, ~~which is to say that~~ i.e., bandwidths of interest are less than 100 ~~Hertz~~ Hz.

The neutral vowel, the plosive /t/, and the fricative /f/ were used as stimuli in experiments in which phonemes served as the standard stimuli. The stimulus pairs were presented at levels approximating normal conversational level speech, i.e., 70, 32, and 35 dB Sensation Level respectively. Discriminability was observed to be strongly dependent on bandwidth and delay. A very slight, if any, decrease in discriminability was observed with increased center frequency. Rank ordering of the phonemes in decreasing sensitivity to dispersion is plosive, vowel, and fricative. For bandwidths of approximately 100 to 400 Hertz, the plosive requires 4 to 8, the vowel 8 to 16, and the fricative 16 to 32 milliseconds of delay. For wider bandwidths, the respective ranges are 2 to 4, 2 to 4, and 4 to 8 milliseconds. More explicitly, for narrower bandwidths, the plosive requires 4 to 8 milliseconds for 100 to 200 Hertz bandwidth, the vowel 8 to 16 milliseconds for 100 to 400 Hertz bandwidth, and the fricative 16 to 32 milliseconds for 100 to 400 Hertz bandwidth. For wider bandwidths the plosive requires 2 to 4 milliseconds for 200 to 400 Hertz bandwidth, the vowel 2 to 4 milliseconds for 200 to 800 Hertz bandwidth, and the fricative 4 to 8 milliseconds for 400 to 800 Hertz bandwidth. Less delay is required for greater bandwidth since discriminability is directly dependent on both independent variables.

Discriminability scores obtained from tests using words and a sentence as standard stimuli agree with the scores obtained from tests using phonemes. Discriminability half way between chance and perfect performance was obtained for the dispersed sentence in which a 400 Hertz band of components centered at 500 Hertz was delayed between 4 and 8 milliseconds. The score agrees with that for a vowel which implies that cues to dispersed sound quality in continuous speech arises from vowel dispersions.

The results of the study indicates that speech processing systems introducing no more than a few milliseconds of dispersion will cause little detriment to the speech quality. This conclusion is significant inasmuch as it is difficult, if not impossible, to determine the phase of many speech processing systems.